



(19) 대한민국특허청(KR)  
(12) 등록특허공보(B1)

(45) 공고일자 2021년08월09일  
(11) 등록번호 10-2287968  
(24) 등록일자 2021년08월03일

(51) 국제특허분류(Int. Cl.)  
A61B 5/00 (2021.01) A61B 5/01 (2021.01)  
A61B 5/021 (2006.01) A61B 5/024 (2006.01)  
A61B 5/08 (2006.01)  
(52) CPC특허분류  
A61B 5/7275 (2013.01)  
A61B 5/01 (2021.01)  
(21) 출원번호 10-2019-0097366  
(22) 출원일자 2019년08월09일  
심사청구일자 2019년08월09일  
(65) 공개번호 10-2020-0114975  
(43) 공개일자 2020년10월07일  
(30) 우선권주장  
1020190036982 2019년03월29일 대한민국(KR)  
(56) 선행기술조사문헌  
Bernard Hernanderz 외 7명, Supervised  
learning for infection risk inference using  
pathology data, BMC Medical Informatics and  
Decision Making (2017) 17:168  
(뒷면에 계속)

(73) 특허권자  
연세대학교 산학협력단  
서울특별시 서대문구 연세로 50 (신촌동, 연세대  
학교)  
주식회사 셀바스에이아이  
서울특별시 금천구 가산디지털1로 19, 20층(가  
산동, 대륭테크노타운18차)  
(72) 발명자  
송영구  
서울특별시 서초구 명달로1길 51, 102동 803호(방  
배동, 방배어울림)  
동재준  
서울특별시 종로구 송월길 99, 203동 904호(홍파  
동, 경희궁자이 2단지)  
(뒷면에 계속)  
(74) 대리인  
특허법인인벤싱크

전체 청구항 수 : 총 24 항

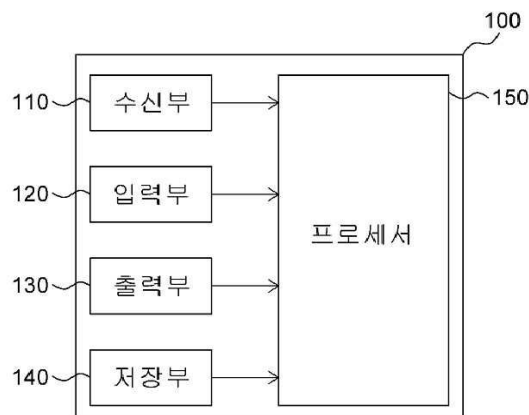
심사관 : 유창용

(54) 발명의 명칭 균혈증 발병 위험도 예측 방법 및 이를 이용한 균혈증 발병 위험도 예측용 디바이스

(57) 요약

본 명세서에서는, 프로세서에 의해 구현되는 균혈증 발병 위험도를 예측하는 방법으로서, 개체에 대한 생물학적 시험 데이터를 수신하는 단계, 생물학적 시험 데이터를 기초로 개체에 대한 균혈증 발병 위험도를 예측하도록 구성된 균혈증 예측 모델을 이용하여 개체의 균혈증 발병 위험도를 예측하는 단계를 포함하고, 생물학적 시험 데이터가 크레아틴 (Creatinine) 수치, 알부민 (Albumin) 수치, CRP (Creactive protein) 수치, WBC (white blood cell) 최소 수치, WBC 최고 수치, 혈소판 (Platelet) 수, 프로트롬빈 (Prothrombin) 시간 중 적어도 하나 및 ALP (Alkaline phosphatase) 수치를 포함하는, 균혈증 발병 위험도의 예측 방법 및 이를 이용한 균혈증 발병 위험도 예측용 디바이스가 제공된다.

대표도 - 도1b



(52) CPC특허분류

**A61B 5/021** (2013.01)  
**A61B 5/024** (2013.01)  
**A61B 5/08** (2013.01)  
**A61B 5/7235** (2021.01)

(72) 발명자

**이경화**

서울특별시 강남구 선릉로69길 19, 111동 1702호(역삼동, 역삼래미안아파트)

**채명훈**

경기도 동두천시 중앙로 41, 303동 1508호(송내동, 송내주공아파트)

**이병수**

경기도 고양시 덕양구 충경로 149, 105동 201호(행신동, 소만마을1단지아파트)

(56) 선행기술조사문헌

Athanasios Tsoukalas 외 2명, From Data to Optimal Decision Making: A Data-Driven, Probabilistic Machine Learning Approach to Decision Support for Patients With Sepsis, JMIR Med Inform 2015;3(1):e11

JP2008522604 A

KR1020170023382 A

JP2012502671 A

Mical Paul, et al., Prediction of Bacteremia Using TREAT, a Computerized Decision-Support System, Clinical Infectious Diseases, Vol.42, Iss.9, 1 May 2006, pp.1274~1282

Cole Beeler, et al., Assessing patient risk of central line-associated bacteremia via machine learning, American Journal of Infection Control, Vol.46, Iss.9, September 2018, pp.986~991

이 발명을 지원한 국가연구개발사업

과제고유번호	HI18C0718
부처명	보건복지부
과제관리(전문)기관명	한국보건산업진흥원
연구사업명	연구자 주도 질병극복연구
연구과제명	중증 감염질환 극복을 위한 인공지능-기반 예측모델
기 여 율	1/1
과제수행기관명	연세대학교 산학협력단
연구기간	2018.04.30 ~ 2019.12.31

공지예외적용 : 있음

## 명세서

### 청구범위

#### 청구항 1

프로세서 및 수신부에 의해 구현되는 균혈증 발병 위험도를 예측하는 방법에 있어서,  
 상기 수신부를 통해, 개체에 대한 생물학적 시험 데이터를 수신하는 단계;  
 상기 프로세서를 통해, 상기 생물학적 시험 데이터를 기초로 상기 개체에 대한 균혈증 발병 위험도를 예측하도록 구성된 균혈증 예측 모델을 이용하여, 상기 개체의 균혈증 발병 위험도를 예측하는 단계를 포함하고,  
 상기 생물학적 시험 데이터는,  
 크레아틴 (Creatinine) 수치, 알부민 (Albumin) 수치, CRP (C-reactive protein) 수치, WBC (white blood cell) 최소 수치, WBC 최고 수치, 프로트롬빈 (Prothrombin) 시간 중 적어도 하나, 혈소판 (Platelet) 수 및 ALP (Alkaline phosphatase) 수치를 포함하고,  
 상기 수신부를 통해, 상기 개체에 대한 생체 신호 데이터를 수신하는 단계를 더 포함하고,  
 상기 균혈증 예측 모델은,  
 상기 생물학적 시험 데이터 및 상기 생체 신호 데이터를 기초로 균혈증 발병 위험도를 예측하도록 더 구성되고,  
 상기 생체 신호 데이터는,  
 SBP (systolic blood pressure), DBP (diastolic blood pressure), 최저 체온, 심박수 및 호흡수 중 적어도 하나, 및 최고 체온을 포함하는, 균혈증 발병 위험도의 예측 방법.

#### 청구항 2

삭제

#### 청구항 3

제1항에 있어서,  
 상기 수신부를 통해, 상기 개체에 대한 임상적 특징 데이터를 수신하는 단계를 더 포함하고,  
 상기 균혈증 예측 모델은,  
 상기 생물학적 시험 데이터 및 상기 임상적 특징 데이터를 기초로 균혈증 발병 위험도를 예측하도록 더 구성되고,  
 상기 임상적 특징 데이터는,  
 성별, 연령, 혈액 배양 기간, ICU (intensive care unit) 치료 여부, 중심 정맥 카테터 (Central venous catheter) 삽입 여부, 스테로이드 치료 여부 및 항생제 치료 여부 중 적어도 하나인, 균혈증 발병 위험도의 예측 방법.

#### 청구항 4

제1항에 있어서,  
 상기 생물학적 시험 데이터는,  
 상기 크레아틴 수치, 상기 알부민 수치, 상기 CRP 수치, 상기 WBC 최소 수치, 상기 WBC 최고 수치, 상기 프로트롬빈 시간 중 적어도 하나, 및 상기 혈소판 수, 상기 ALP 수치를 포함하는, 균혈증 발병 위험도의 예측 방법.

#### 청구항 5

제1항에 있어서,

상기 생물학적 시험 데이터는,

ALT (alanine aminotransferase) 수치, 트롬보플라스틴 (thromboplastin) 의 응고 시간, AST (aspartate aminotransferase) 수치, CRP (C-reactive protein) 수치, 적혈구 침강 속도, FERR (ferritin) 수치, HMG (hemoglobin) 수치, PTINR (Prothrombin Time International Normalized Ratio), PTPER (Prothrombin time (%)), PTSEC (Prothrombin time (sec)), 적혈구 수치, TBIL (Total bilirubin) 수치, TCO<sub>2</sub> (total carbon dioxide) 수치 및 백혈구 수치 중 적어도 하나를 더 포함하는, 균혈증 발병 위험도의 예측 방법.

#### 청구항 6

제1항에 있어서,

상기 균혈증은, 급성 중증 균혈증 (acute severe bacteraemia) 인, 균혈증 발병 위험도의 예측 방법.

#### 청구항 7

제1항에 있어서,

상기 균혈증 예측 모델은,

MLP (multi-layer perceptron), SVM (support vector machine), RF (random forest), 및 복수의 모델의 조합으로 구성된 앙상블 모델, 선행 학습된 (pre-trained) 모델, Xgboost 모델 중 적어도 하나의 모델인, 균혈증 발병 위험도의 예측 방법.

#### 청구항 8

제7항에 있어서,

상기 균혈증 예측 모델은,

상기 MLP 모델 또는 상기 Xgboost 모델이고,

상기 MLP 모델은 128 노드 (nodes) 또는 256 노드를 갖는 단일의 히든 레이어 (hidden layer) 를 포함하거나,

상기 단일의 히든 레이어가 복수로 존재하는 복수의 히든 레이어를 포함하는, 균혈증 발병 위험도의 예측 방법.

#### 청구항 9

제1항에 있어서,

상기 균혈증 예측 모델은,

상기 수신부를 통해, 균혈증 발병 표본 개체 및 정상 표본 개체에 대한 생물학적 실험 데이터, 생체 신호 데이터 및 임상적 특징 데이터로 구성된 학습용 데이터를 수신하는 단계, 및

상기 프로세서를 통해, 상기 학습용 데이터를 기초로, 균혈증 또는 정상을 예측하는 단계를 통해 학습된 모델인, 균혈증 발병 위험도의 예측 방법.

#### 청구항 10

제9항에 있어서,

상기 학습용 데이터를 수신하는 단계 이후에 수행되는,

상기 학습용 데이터를 평가하는 단계를 더 포함하고,

상기 학습용 데이터를 평가하는 단계는,

상기 프로세서를 통해, 상기 학습용 데이터에 대하여 균혈증과의 관련도 (relevance) 점수를 산출하는 단계, 및

상기 관련도 점수를 기초로, 미리 결정된 순위 내에 있는 균혈증 관련 학습용 데이터를 결정하는 단계를 포함하는, 균혈증 발병 위험도의 예측 방법.

#### 청구항 11

제10항에 있어서,

상기 학습용 데이터를 평가하는 단계는,

상기 프로세서를 통해, OOS (one-out search), LRP (Layer-wise Relevance Propagation), GI (gradient input), IG (integrated gradients), 및 SM (saliency map) 중 적어도 하나의 알고리즘에 기초하여,

상기 학습용 데이터에 대하여 균혈증과의 관련도 점수를 산출하는 단계, 및

상기 관련도 점수를 기초로, 미리 결정된 순위 내에 있는 균혈증 관련 학습용 데이터를 결정하는 단계를 포함하는, 균혈증 발병 위험도의 예측 방법.

#### 청구항 12

제1항에 있어서,

상기 프로세서를 통해, 상기 개체에 대하여 예측된 상기 균혈증 발병 위험도를 제공하는 단계를 더 포함하는, 균혈증 발병 위험도의 예측 방법.

#### 청구항 13

제12항에 있어서,

상기 균혈증 발병 위험도를 제공하는 단계는,

상기 균혈증 예측 모델에 의해 개체에 대한 균혈증 발병 위험도가 예측될 경우,

상기 프로세서를 통해, 위험 알림을 제공하는 단계를 포함하는, 균혈증 발병 위험도의 예측 방법.

#### 청구항 14

제12항에 있어서,

상기 균혈증 발병 위험도를 제공하는 단계는,

상기 프로세서를 통해, 상기 균혈증 예측 모델에 의해 개체에 대한 균혈증 발병 위험 확률을 산출하는 단계, 및

상기 균혈증 발병 위험 확률을 제공하는 단계를 포함하는, 균혈증 발병 위험도의 예측 방법.

#### 청구항 15

개체에 대한 생물학적 시험 데이터를 수신하도록 구성된 수신부, 및

상기 수신부와 연결된 프로세서를 포함하고,

상기 프로세서는, 상기 생물학적 시험 데이터를 기초로 상기 개체에 대한 균혈증 발병 위험도를 예측하도록 구성된 균혈증 예측 모델을 이용하여, 상기 개체의 균혈증 발병 위험도를 예측하도록 구성되고,

상기 생물학적 시험 데이터는,

크레아틴 (Creatinine) 수치, 알부민 (Albumin) 수치, CRP (C-reactive protein) 수치, WBC (white blood cell) 최소 수치, WBC 최고 수치, 프로트롬빈 (Prothrombin) 시간 중 적어도 하나, 혈소판 (Platelet) 수 및 ALP (Alkaline phosphatase) 수치를 포함하고,

상기 수신부는, 개체에 대한 생체 신호 데이터를 수신하도록 더 구성되고,

상기 균혈증 예측 모델은,

상기 생물학적 시험 데이터 및 상기 생체 신호 데이터를 기초로 균혈증 발병 위험도를 예측하도록 더 구성되고,

상기 생체 신호 데이터는,

SBP (systolic blood pressure), DBP (diastolic blood pressure), 최저 체온, 심박수 및 호흡수 중 적어도 하나, 및 최고 체온을 포함하는, 균혈증 발병 위험도의 예측용 디바이스.

#### 청구항 16

삭제

#### 청구항 17

제15항에 있어서,

상기 수신부는 상기 개체에 대한 임상적 특징 데이터를 수신하도록 더 구성되고,

상기 균혈증 예측 모델은,

상기 생물학적 시험 데이터 및 상기 임상적 특징 데이터를 기초로 균혈증 발병 위험도를 예측하도록 더 구성되고,

상기 임상적 특징 데이터는,

성별, 연령, 혈액 배양 기간, ICU (intensive care unit) 치료 여부, 중심 정맥 카테터 (Central venous catheter) 삽입 여부, 스테로이드 치료 여부 및 항생제 치료 여부 중 적어도 하나인, 균혈증 발병 위험도의 예측용 디바이스.

#### 청구항 18

제15항에 있어서,

상기 생물학적 시험 데이터는,

상기 크레아틴 수치, 상기 알부민 수치, 상기 CRP 수치, 상기 WBC 최소 수치, 상기 WBC 최고 수치, 상기 프로트롬빈 시간 중 적어도 하나, 및 상기 혈소판 수, 상기 ALP 수치를 포함하는, 균혈증 발병 위험도의 예측용 디바이스.

#### 청구항 19

제15항에 있어서,

상기 생물학적 시험 데이터는,

ALT (alanine aminotransferase) 수치, 트롬보플라스틴 (thromboplastin) 의 응고 시간, AST (aspartate aminotransferase) 수치, CRP (C-reactive protein) 수치, 적혈구 침강 속도, FERR (ferritin) 수치, HMG (hemoglobin) 수치, PTINR (Prothrombin Time International Normalized Ratio), PTPER (Prothrombin time (%)), PTSEC (Prothrombin time (sec)), 적혈구 수치, TBIL (Total bilirubin) 수치, TCO<sub>2</sub> (total carbon dioxide) 수치 및 백혈구 수치 중 적어도 하나를 더 포함하는, 균혈증 발병 위험도의 예측용 디바이스.

#### 청구항 20

제15항에 있어서,

상기 균혈증은, 급성 중증 균혈증 (acute severe bacteraemia) 인, 균혈증 발병 위험도의 예측용 디바이스.

#### 청구항 21

제15항에 있어서,

상기 균혈증 예측 모델은,

MLP (multi-layer perceptron), SVM (support vector machine), RF (random forest), 및 복수의 모델의 조합으로 구성된 앙상블 모델, 선행 학습된 (pre-trained) 모델, Xgboost 모델 중 적어도 하나의 모델인, 균혈증 발병 위험도의 예측용 디바이스.

#### 청구항 22

제21항에 있어서,

상기 균혈증 예측 모델은,

상기 MLP 모델 또는 상기 Xgboost 모델이고,

상기 MLP 모델은 128 노드 (nodes) 또는 256 노드를 갖는 단일의 히든 레이어 (hidden layer) 를 포함하거나,

상기 단일의 히든 레이어가 복수로 존재하는 복수의 히든 레이어를 포함하는, 균혈증 발병 위험도의 예측용 디바이스.

### 청구항 23

제15항에 있어서,

상기 균혈증 예측 모델은,

균혈증 발병 표본 개체 및 정상의 표본 개체에 대한 생물학적 실험 데이터, 생체 신호 데이터 및 임상적 특징 데이터로 구성된 학습용 데이터를 수신하고, 상기 학습용 데이터를 기초로, 균혈증 또는 정상을 예측하는 단계를 통해 학습된 모델인, 균혈증 발병 위험도의 예측용 디바이스.

### 청구항 24

제15항에 있어서,

상기 개체에 대하여 예측된 상기 균혈증 발병 위험도를 제공하도록 구성된 출력부를 더 포함하는, 균혈증 발병 위험도의 예측용 디바이스.

### 청구항 25

제24항에 있어서,

상기 출력부는,

상기 균혈증 예측 모델에 의해 개체에 대한 균혈증 발병 위험도가 예측될 경우,

위험 알람을 제공하도록 구성된, 균혈증 발병 위험도의 예측용 디바이스.

### 청구항 26

제24항에 있어서,

상기 균혈증 예측 모델은,

상기 개체에 대한 균혈증 발병 위험 확률을 산출하도록 더 구성되고,

상기 출력부는,

상기 균혈증 발병 위험 확률을 제공하도록 더 구성된, 균혈증 발병 위험도의 예측용 디바이스.

## 발명의 설명

### 기술 분야

[0001] 본 발명은 균혈증 발병 위험도의 예측 방법 및 디바이스에 관한 것으로, 보다 구체적으로 개체로부터 획득한 다양한 임상 데이터를 기초로 균혈증 발병의 위험도를 예측하고 제공하도록 구성된 방법 및 이를 이용한 디바이스에 관한 것이다.

### 배경 기술

[0002] 균혈증은 균이 혈액 속에 들어가서 온몸을 순환하고 있는 임상적 상태를 지칭할 수 있다. 이러한 균혈증은 세균이 세망 내피계의 제거 능력을 넘는 속도로 증식할 때 발생할 수 있다.

[0003] 이때, 균혈증은 임상적 양상에 따라, 일시적 균혈증과 간헐적 균혈증, 지속적 균혈증으로 구분될 수 있다. 보다 구체적으로, 일시적 균혈증은 전신성 또는 국소적 감염 초기에 발생하는 균혈증으로, 뇌막염, 폐렴, 화농성 관절염, 골수염, 임균 또는 수막구균 감염을 동반할 수 있다. 또한, 간헐적 균혈증은 복강, 골반강, 신장주위,

간, 전립선 등의 부위에 농양이 있을 때 발병될 수 있고, 지속적 균혈증은 세균성 심내막염과 혈관 내 카테테의 감염이 있을 때 발병될 수 있고 장티푸스, 부르셀라증을 동반할 수 있다.

[0004] 한편, 균혈증의 치료 방법으로는 항생제의 사용이 있을 수 있으며 초기에 적절한 항생제를 경험적으로 사용하는 것이 균혈증에서 패혈증, 패혈성 쇼크로 진행되는 것을 막을 수 있다.

[0005] 따라서, 균혈증의 발병 전 균혈증을 예측하는 것은 임상적으로 매우 중요할 수 있다.

[0006] 이에, 균혈증이 발생하기 전에 이를 미리 예측하고 빠른 조치를 취할 수 있는, 새로운 균혈증 발병 위험도 예측 시스템의 개발이 지속적으로 요구되고 있는 실정이다.

[0007] 발명의 배경이 되는 기술은 본 발명에 대한 이해를 보다 용이하게 하기 위해 작성되었다. 발명의 배경이 되는 기술에 기재된 사항들이 선행기술로 존재한다고 인정하는 것으로 이해되어서는 안 된다.

선행문헌: Bernard Hernandez 외 7명, Supervised learning for infection risk inference using pathology data, BMC Medical Informatics and Decision Making (2017) 17:168

## 발명의 내용

### 해결하려는 과제

[0008] 한편, 본 발명의 발명자들은, 임상적으로 균혈증과 같은 어떠한 질환의 이환이 의심되는 상황이 도래되기 이전에 미리 인체의 생리학적 대응 반응의 일환으로 생체 신호들의 변화가 선행할 것이라는 점에 주목하였다.

[0009] 보다 구체적으로, 본 발명의 발명자들은, 일반 병실 또는 중환자실에서 획득할 수 있는 다양한 임상 데이터의 변화가 균혈증과 같은 잠재적 질환과 연관성을 가질 수 있음에 주목하였다.

[0010] 특히, 본 발명의 발명자들은 균혈증이 외부적 환경에 의해 2 차적으로 야기될 수 있는 질환임에 주목하였고, 환자로부터 분리된 혈액과 같은 생물학적 시료의 분석 결과의 생물학적 시험 데이터뿐만 아니라, 다양한 생체 신호 데이터, 의료적 처치와 관련한 데이터들이 균혈증의 발병과 연관이 있음을 인지할 수 있었다.

[0011] 나아가, 본 발명의 발명자들은, 혈액 속에서 균이 증식하여 발열, 저혈압 등의 중독 증상을 일으키는 패혈증과 구별되는, 균혈증의 예측의 정확도를 높이는 생물학적 특징을 탐색하고자 하였다.

[0012] 그 결과, 본 발명의 발명자들은 개체로부터 획득한 생물학적 시험 데이터들 중 특히 ALP (Alkaline phosphatase) 수치와 함께 혈소판 수치가 균혈증의 발병과 연관이 있음을 발견할 수 있었다.

[0013] 이에, 본 발명의 발명자들은, ALP (Alkaline phosphatase) 수치 및/또는 혈소판 수치를 포함하는 생물학적 시험 데이터와 함께 생체 신호 데이터, 임상적 특징 데이터들을 더욱 고려하여 균혈증 발병 위험도를 높은 신뢰도 및 정확도로 예측할 수 있는 균혈증 발병 위험도 예측 시스템을 개발하기에 이르렀다.

[0014] 한편, 본 발명의 발명자들은, 새로운 균혈증 발병 위험도 예측 시스템에 생물학적 시험 데이터, 나아가 생체 신호 데이터, 임상적 특징 데이터들을 기초로 균혈증을 예측하도록 학습된 예측 모델을 적용할 수 있었다.

[0015] 관련하여, 본 발명의 발명자들은, 다양한 임상 데이터 중 균혈증임을 예측하는 것에 영향력이 큰 임상 데이터 또는, 균혈증이 아닌 정상을 예측하는 것에 영향력이 큰 임상 데이터를 결정하도록 평가를 수행하였고, 이를 예측 모델의 학습에 적용하고자 하였다. 이를 통해, 본 발명의 발명자들은 예측 모델의 균혈증 예측의 진단 능력 향상을 기대할 수 있었다.

[0016] 결과적으로, 균혈증 예측 모델에 기초한 새로운 균혈증 발병 위험도 예측 시스템은, 개체에 대하여 정확도 및 신뢰도 높은 균혈증 진단 정보를 제공할 수 있었다.

[0017] 나아가, 본 발명의 발명자들은 이러한 균혈증 발병 위험도 예측 시스템 개발을 통해, 환자의 상태에 대하여 빠르게 감지하고 균혈증 발병 위험을 미리 인지하여, 환자에 대한 처치 시점을 앞당겨 균혈증의 치료 성과를 높일 수 있음을 기대할 수 있었다. 나아가, 본 발명의 발명자들은 균혈증 발병 위험도의 예측 시스템의 개발이 환자의 생존률 증가, 합병증 예방, 및 치료 비용의 감소 효과를 제공할 수 있다는 것을 더욱 기대할 수 있었다.

[0018] 또한, 본 발명의 발명자들은, 보호자 또는 의료진이 중환자와 같은 지속적인 모니터링이 요구되는 개체에 대하여 보다 용이하게 균혈증 발병 고 위험군을 인지할 수 있도록, 개체에 대하여 예측 모델에 의해 균혈증이 예측될 경우, 알람을 제공하여 균혈증 발병 위험도를 알리도록 구성하였다.



- [0019] 이에, 본 발명이 해결하고자 하는 과제는, 수신된 개체에 대한 생물학적 시험 데이터, 나아가 생체 신호 데이터, 임상적 특징 데이터를 기초로, 균혈증 발병 위험도를 예측하도록 구성된 균혈증 예측 모델을 이용하여, 개체에 대한 균혈증 발병 위험도를 예측하도록 구성된, 균혈증 발병 위험도 예측 방법을 제공하는 것이다.
- [0020] 본 발명이 해결하고자 하는 또 다른 과제는, 개체에 대한 생물학적 시험 데이터, 나아가 생체 신호 데이터, 임상적 특징 데이터를 수신하도록 구성된 수신부, 및 수신부와 통신하도록 연결되고 예측 모델에 기초하여 개체에 균혈증 발병 위험도를 예측하도록 구성된 프로세서를 포함하는, 균혈증 발병 위험도 예측용 디바이스를 제공하는 것이다.
- [0021] 본 발명의 과제들은 이상에서 언급한 과제들로 제한되지 않으며, 언급되지 않은 또 다른 과제들은 아래의 기재로부터 당업자에게 명확하게 이해될 수 있을 것이다.

### 과제의 해결 수단

- [0022] 전술한 바와 같은 과제를 해결하기 위해, 본 발명의 일 실시예에 따른 균혈증 발병 위험도의 예측 방법이 제공된다. 이때, 본 발명의 균혈증 예측 방법은, 프로세서에 의해 구현되는 균혈증 발병 위험도를 예측하도록 구성된 것으로, 개체에 대한 생물학적 시험 데이터를 수신하는 단계, 생물학적 시험 데이터를 기초로 개체에 대한 균혈증 발병 위험도를 예측하도록 구성된 균혈증 예측 모델을 이용하여, 개체의 균혈증 발병 위험도를 예측하는 단계를 포함한다.
- [0023] 이때 생물학적 시험 데이터는, 크레아틴 (Creatinine) 수치, 알부민 (Albumin) 수치, CRP (C-reactive protein) 수치, WBC (white blood cell) 최소 수치, WBC 최고 수치, 혈소판 (Platelet) 수, 프로트롬빈 (Prothrombin) 시간 및 ALP (Alkaline phosphatase) 수치를 포함한다.
- [0024] 본 발명의 특징에 따르면 본 발명의 균혈증 예측 방법은, 개체에 대한 생체 신호 데이터를 수신하는 단계를 더 포함할 수 있고, 균혈증 예측 모델은 생물학적 시험 데이터 및 생체 신호 데이터를 기초로 균혈증 발병 위험도를 예측하도록 더 구성될 수 있다. 이때, 생체 신호 데이터는, SBP (systolic blood pressure), DBP (diastolic blood pressure), 최고 체온, 최저 체온, 심박수 및 호흡수 중 적어도 하나일 수 있다.
- [0025] 본 발명의 다른 특징에 따르면 본 발명의 균혈증 예측 방법은, 개체에 대한 임상적 특징 데이터를 수신하는 단계를 더 포함하고, 균혈증 예측 모델은 생물학적 시험 데이터 및 임상적 특징 데이터를 기초로 균혈증 발병 위험도를 예측하도록 더 구성될 수 있다. 이때, 임상적 특징 데이터는, 성별, 연령, 혈액 배양 기간, ICU (intensive care unit) 치료 여부, 중심 정맥 카테터 (Central venous catheter) 삽입 여부, 스테로이드 치료 여부 및 항생제 치료 여부 중 적어도 하나일 수 있다.
- [0026] 본 발명의 또 다른 특징에 따르면, 생물학적 시험 데이터는, 크레아틴 수치, 알부민 수치, CRP 수치, WBC 최소 수치, WBC 최고 수치, 프로트롬빈 시간, 및 혈소판 수, ALP 수치를 포함할 수 있다.
- [0027] 본 발명의 또 다른 특징에 따르면, 생물학적 시험 데이터는 ALT (alanine aminotransferase) 수치, 트롬보플라스틴 (thromboplastin) 의 응고 시간, AST (aspartate aminotransferase) 수치, CRP (C-reactive protein) 수치, 적혈구 침강 속도, FERR (ferritin) 수치, HMG (hemoglobin) 수치, PTINR (Prothrombin Time International Normalized Ratio), PTPER (Prothrombin Time (%)), PTSEC (Prothrombin Time (sec)), 적혈구 수치, TBIL (Total bilirubin) 수치, TCO<sub>2</sub> (total carbon dioxide) 수치 및 백혈구 수치 중 적어도 하나를 더 포함
- [0028] 본 발명의 또 다른 특징에 따르면, 균혈증은, 급성 중증 균혈증 (acute severe bacteraemia) 일 수 있다.
- [0029] 본 발명의 또 다른 특징에 따르면, 균혈증 예측 모델은, MLP (multi-layer perceptron), SVM (support vector machine), RF (random forest), 및 복수의 모델의 조합으로 구성된 앙상블 모델, 선행 학습된 (pre-trained) 모델, Xgboost 모델 중 적어도 하나의 모델일 수 있다.
- [0030] 본 발명의 또 다른 특징에 따르면, 균혈증 예측 모델은 또는 Xgboost 모델일 수 있다. 이때, MLP 모델은 128 노드 (nodes) 또는 256 노드를 갖는 단일의 히든 레이어 (hidden layer) 를 포함하거나, 단일의 히든 레이어가 복수로 존재하는 복수의 히든 레이어를 포함할 수 있다.
- [0031] 본 발명의 또 다른 특징에 따르면, 균혈증 예측 모델은, 균혈증 발병 표본 개체 및 정상 표본 개체에 대한 생물학적 실험 데이터, 생체 신호 데이터 및 임상적 특징 데이터로 구성된 학습용 데이터를 수신하는 단계, 및 학습용 데이터를 기초로, 균혈증 또는 정상을 예측하는 단계를 통해 학습된 모델일 수 있다. 이때, 정상 표본

개체는, 임상적으로 균혈증 발생되지 않고 균혈증이 아닌 것으로 평가된 개체일 수 있다.

- [0032] 본 발명의 또 다른 특징에 따르면, 본 발명의 균혈증 예측 방법은, 학습용 데이터를 수신하는 단계 이후에 수행되는 학습용 데이터를 평가하는 단계를 더 포함할 수 있다. 이때, 학습용 데이터를 평가하는 단계는, 학습용 데이터에 대하여 균혈증과의 관련도 (relevance) 점수를 산출하는 단계, 및 관련도 점수를 기초로, 미리 결정된 순위 내에 있는 균혈증 관련 학습용 데이터를 결정하는 단계를 포함할 수 있다.
- [0033] 본 발명의 또 다른 특징에 따르면, 학습용 데이터를 평가하는 단계는, OOS (one-out search), LRP (Layer-wise Relevance Propagation), GI (gradient input), IG (integrated gradients), 및 SM (saliency map) 중 적어도 하나의 알고리즘에 기초하여, 학습용 데이터에 대하여 균혈증과의 관련도 점수를 산출하는 단계, 및 관련도 점수를 기초로, 미리 결정된 순위 내에 있는 균혈증 관련 학습용 데이터를 결정하는 단계를 포함할 수 있다.
- [0034] 본 발명의 또 다른 특징에 따르면, 본 발명의 균혈증 예측 방법은, 개체에 대하여 예측된 균혈증 발병 위험도를 제공하는 단계를 더 포함할 수 있다.
- [0035] 본 발명의 또 다른 특징에 따르면, 균혈증 발병 위험도를 제공하는 단계는, 균혈증 예측 모델에 의해 개체에 대한 균혈증 발병 위험도가 예측될 경우, 위험 알림을 제공하는 단계를 포함할 수 있다.
- [0036] 본 발명의 또 다른 특징에 따르면, 균혈증 발병 위험도를 제공하는 단계는, 균혈증 예측 모델에 의해 개체에 대한 균혈증 발병 위험 확률을 산출하는 단계, 및 균혈증 발병 위험 확률을 제공하는 단계를 포함할 수 있다.
- [0037] 진술한 바와 같은 과제를 해결하기 위해, 본 발명의 다른 실시예에 따른 균혈증 발병 위험도의 예측용 디바이스가 제공된다. 이때, 본 발명의 균혈증 예측용 디바이스는, 개체에 대한 생물학적 시험 데이터를 수신하도록 구성된 수신부, 및 수신부와 연결된 프로세서를 포함한다. 이때, 프로세서는, 생물학적 시험 데이터를 기초로 개체에 대한 균혈증 발병 위험도를 예측하도록 구성된 균혈증 예측 모델을 이용하여, 개체의 균혈증 발병 위험도를 예측하도록 구성된다. 나아가, 생물학적 시험 데이터는, 크레아틴 (Creatinine) 수치, 알부민 (Albumin) 수치, CRP (C-reactive protein) 수치, WBC (white blood cell) 최소 수치, WBC 최고 수치, 혈소판 (Platelet) 수, 프로트롬빈 (Prothrombin) 시간 및 ALP (Alkaline phosphatase) 수치를 포함한다.
- [0038] 본 발명의 특징에 따르면, 수신부는, 개체에 대한 생체 신호 데이터를 수신하도록 더 구성되고, 균혈증 예측 모델은, 생물학적 시험 데이터 및 생체 신호 데이터를 기초로 균혈증 발병 위험도를 예측하도록 더 구성될 수 있다. 이때, 생체 신호 데이터는, SBP (systolic blood pressure), DBP (diastolic blood pressure), 최고 체온, 최저 체온, 심박수 및 호흡수 중 적어도 하나일 수 있다.
- [0039] 본 발명의 다른 특징에 따르면, 수신부는 개체에 대한 임상적 특징 데이터를 수신하도록 더 구성되고, 균혈증 예측 모델은 생물학적 시험 데이터 및 임상적 특징 데이터를 기초로 균혈증 발병 위험도를 예측하도록 더 구성될 수 있다. 이때, 임상적 특징 데이터는, 성별, 연령, 혈액 배양 기간, ICU (intensive care unit) 치료 여부, 중심 정맥 카테터 (Central venous catheter) 삽입 여부, 스테로이드 치료 여부 및 항생제 치료 여부 중 적어도 하나일 수 있다.
- [0040] 본 발명의 또 다른 특징에 따르면, 생물학적 시험 데이터는, 크레아틴 수치, 알부민 수치, CRP 수치, WBC 최소 수치, WBC 최고 수치, 프로트롬빈 시간, 및 혈소판 수, ALP 수치를 포함할 수 있다.
- [0041] 본 발명의 또 다른 특징에 따르면, 생물학적 시험 데이터는, ALT 수치, 트롬보플라스틴의 응고 시간, AST 수치, CRP 수치, 적혈구 침강 속도, 페리틴 (Ferritin) 수치, 헤모글로빈 수치, PTINR, PTPER, PTSEC, 적혈구 수치, 총 빌리루빈 수치, TCO<sub>2</sub> 수치 및 백혈구 수치 중 적어도 하나를 더 포함할 수 있다.
- [0042] 본 발명의 또 다른 특징에 따르면, 균혈증은, 급성 중증 균혈증 (acute severe bacteraemia) 일 수 있다.
- [0043] 본 발명의 또 다른 특징에 따르면, 균혈증 예측 모델은, MLP (multi-layer perceptron), SVM (support vector machine), RF (random forest), 및 복수의 모델의 조합으로 구성된 앙상블 모델, 선행 학습된 (pre-trained) 모델, Xgboost 모델 중 적어도 하나의 모델일 수 있다.
- [0044] 본 발명의 또 다른 특징에 따르면, 균혈증 예측 모델은 MLP 모델 또는 Xgboost 모델일 수 있다. 이때, MLP 모델은 128 노드 (nodes) 또는 256 노드를 갖는 단일의 히든 레이어 (hidden layer) 를 포함하거나, 단일의 히든 레이어가 복수로 존재하는 복수의 히든 레이어를 포함할 수 있다.
- [0045] 본 발명의 또 다른 특징에 따르면, 균혈증 예측 모델은, 균혈증 발병 표본 개체 및 정상의 표본 개체에 대한 생물학적 실험 데이터, 생체 신호 데이터 및 임상적 특징 데이터로 구성된 학습용 데이터를 수신하고, 학습용 데

이터를 기초로, 균혈증 또는 정상을 예측하는 단계를 통해 학습된 모델일 수 있다. 이때, 정상의 표본 개체는, 임상적으로 균혈증 발생되지 않고 균혈증이 아닌 것으로 평가된 개체일 수 있다.

[0046] 본 발명의 또 다른 특징에 따르면, 개체에 대하여 예측된 균혈증 발병 위험도를 제공하도록 구성된 출력부를 더 포함할 수 있다.

[0047] 본 발명의 또 다른 특징에 따르면, 출력부는, 균혈증 예측 모델에 의해 개체에 대한 균혈증 발병 위험도가 예측될 경우, 위험 알림을 제공하도록 구성될 수 있다.

[0048] 본 발명의 또 다른 특징에 따르면, 균혈증 예측 모델은 개체에 대한 균혈증 발병 위험 확률을 산출하도록 더 구성될 수 있다. 나아가, 출력부는, 균혈증 발병 위험 확률을 제공하도록 더 구성될 수 있다.

[0049] 이하, 실시예를 통하여 본 발명을 보다 상세히 설명한다. 다만, 이들 실시예는 본 발명을 예시적으로 설명하기 위한 것에 불과하므로 본 발명의 범위가 이들 실시예에 의해 한정되는 것으로 해석되어서는 아니 된다.

### 발명의 효과

[0050] 본 발명은 개체에 대하여 약생물학적 시험 데이터, 나아가 생체 신호 데이터, 임상적 특징 데이터를 수신하고, 수신된 데이터 기반의 균혈증 발병 위험도 예측 시스템을 제공함으로써, 개체의 균혈증 발병 위험도에 대하여 빠르게 감지하고, 이와 연관된 정보를 제공할 수 있다.

[0051] 또한, 본 발명은, 생물학적 시험 데이터, 나아가 생체 신호 데이터, 임상적 특징 데이터들을 기초로 균혈증을 예측하도록 학습된 예측 모델을 적용한 예측 시스템을 제공함으로써, 정확도 및 신뢰도 높은 균혈증 발병의 예측 정보를 제공할 수 있다.

[0052] 보다 구체적으로, 본 발명은, 다양한 임상 데이터 중 균혈증임을 예측하는 것에 영향력이 큰 임상 데이터 또는, 균혈증이 아닌 정상 상태를 예측하는 것에 영향력이 큰 임상 데이터를 결정하도록 평가를 수행하였고, 이를 예측 모델의 학습에 적용함으로써 균혈증 예측의 진단 능력 향상된 예측 시스템을 제공할 수 있다.

[0053] 이에, 본 발명은, 패혈증과 구별되는 또는 패혈증과는 상이한 균혈증에 대하여 높은 신뢰도 및 정확도로 발병의 위험도를 예측할 수 있는 효과가 있다.

[0054] 또한, 본 발명은, 개체에 대하여 예측 모델에 의해 균혈증이 예측될 경우, 알림을 제공하여 균혈증 발병 위험도를 알리도록 구성됨에 따라, 보호자 또는 의료진은 중환자와 같은 지속적인 모니터링이 요구되는 개체에 대하여 보다 용이하게 균혈증 발병 고 위험군을 인지할 수 있다.

[0055] 따라서, 본 발명은 균혈증에 대한 처치 시점을 앞당겨 치료에 대한 좋은 예후를 제공할 수 있다.

[0056] 나아가, 본 발명은 균혈증 발병 개체의 생존률 증가, 합병증 예방, 및 치료 비용의 감소 효과를 제공할 수 있다.

[0057] 본 발명에 따른 효과는 이상에서 예시된 내용에 의해 제한되지 않으며, 더욱 다양한 효과들이 본 명세서 내에 포함되어 있다.

### 도면의 간단한 설명

[0058] 도 1a는 본 발명의 일 실시예에 따른 균혈증 발병 위험도의 예측 방법 및 디바이스에 기초한 균혈증 발병 위험도 예측 시스템을 예시적으로 도시한 것이다.

도 1b는 본 발명의 일 실시예에 따른 균혈증 발병 위험도의 예측용 디바이스의 구성을 예시적으로 도시한 것이다.

도 2는 본 발명의 일 실시예에 따른 균혈증 발병 위험도의 예측 방법의 절차를 예시적으로 도시한 것이다.

도 3a 및 3b는 본 발명의 다양한 실시예에 이용되는 균혈증 예측 모델의 학습 데이터를 예시적으로 도시한 것이다.

도 3c는 본 발명의 다양한 실시예에 이용되는 균혈증 예측 모델의 구성을 예시적으로 도시한 것이다.

도 3d는 본 발명의 다양한 실시예에 이용되는 균혈증 예측 모델의 평가 방법을 예시적으로 도시한 것이다.

도 4a 및 4b는 본 발명의 다양한 실시예에 이용되는 균혈증 예측 모델에 대한 제1 평가 결과를 도시한 것이다.

도 4c 및 4d는 본 발명의 다양한 실시예에 이용되는 균혈증 예측 모델에 대한, 균혈증 예측의 영향력에 따른 임상 데이터들을 도시한 것이다.

도 4e 및 4f는 본 발명의 다양한 실시예에 이용되는 균혈증 예측 모델에 대한, 균혈증 예측의 영향력에 따른 임상 데이터들의 변화에 따른 성능 변화를 도시한 것이다.

도 5a 및 5b는 본 발명의 다양한 실시예에 이용되는 균혈증 예측 모델의 제2 평가를 위한 학습 데이터 및 평가 데이터를 도시한 것이다.

도 5c 내지 5i는 본 발명의 다양한 실시예에 이용되는 균혈증 예측 모델에 대한 제2 평가 결과를 도시한 것이다.

### 발명을 실시하기 위한 구체적인 내용

- [0059] 본 발명의 이점 및 특징, 그리고 그것들을 달성하는 방법은 첨부되는 도면과 함께 상세하게 후술되어 있는 실시예들을 참조하면 명확해질 것이다. 그러나, 본 발명은 이하에서 개시되는 실시예들에 한정되는 것이 아니라 서로 다른 다양한 형태로 구현될 것이며, 단지 본 실시예들은 본 발명의 개시가 완전하도록 하며, 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자에게 발명의 범주를 완전하게 알려주기 위해 제공되는 것이며, 본 발명은 청구항의 범주에 의해 정의될 뿐이다.
- [0060] 본 발명의 실시예를 설명하기 위한 도면에 개시된 형상, 크기, 비율, 각도, 개수 등은 예시적인 것이므로 본 발명이 도시된 사항에 한정되는 것은 아니다. 또한, 본 발명을 설명함에 있어서, 관련된 공지 기술에 대한 구체적인 설명이 본 발명의 요지를 불필요하게 흐릴 수 있다고 판단되는 경우 그 상세한 설명은 생략한다. 본 명세서 상에서 언급된 '포함한다', '갖는다', '이루어진다' 등이 사용되는 경우, '~만'이 사용되지 않는 이상 다른 부분이 추가될 수 있다. 구성요소를 단수로 표현한 경우에 특별히 명시적인 기재 사항이 없는 한 복수를 포함하는 경우를 포함한다.
- [0061] 구성요소를 해석함에 있어서, 별도의 명시적 기재가 없더라도 오차 범위를 포함하는 것으로 해석한다.
- [0062] 본 발명의 여러 실시예들의 각각 특징들이 부분적으로 또는 전체적으로 서로 결합 또는 조합 가능하며, 당업자가 충분히 이해할 수 있듯이 기술적으로 다양한 연동 및 구동이 가능하며, 각 실시예들이 서로에 대하여 독립적으로 실시 가능할 수도 있고 연관 관계로 함께 실시 가능할 수도 있다.
- [0063] 본 명세서의 해석의 명확함을 위해, 이하에서는 본 명세서에서 사용되는 용어들을 정의하기로 한다.
- [0064] 본 명세서에서 사용되는 용어, "균혈증"은 혈류 속에 병원균이 살아서 순환하고 있는 임상적 상태를 의미한다. 이때, 균혈증은, 혈액 속에서 균이 증식하여 발열, 저혈압등의 중독 증상을 일으키는 '패혈증'과 상이하며 본 명세서에서는 패혈증과 균혈증을 구분된다.
- [0065] 본 명세서에서 사용되는 용어, "개체"는 균혈증의 발병 위험도를 예측하고자 하는 대상을 의미할 수 있다. 한편, 본원 명세서 내에서 개체는 '환자' 또는 '중환자'일 수 있으나, 이에 제한되는 것은 아니며 균혈증 발병의 위험도를 예측하고자 하는 모든 대상을 포함할 수 있다.
- [0066] 이때, 균혈증 발병 개체는 바실러스 종 (*Bacillus spp.*), 비리단스 그룹 스트렙토코커스 (*Viridans group streptococci*), CNS (coagulase-negative staphylococci) 등 샘플이 오염되는 경우 주로 나타나는 균들이 혈류에 존재하는 경우의 개체와는 상이할 수 있다.
- [0067] 본 명세서에서 사용되는 용어, "생물학적 시험 데이터"는 개체로부터 분리된 생물학적 시료, 예를 들어 혈액에 대한 임상 데이터를 의미할 수 있다.
- [0068] 보다 구체적으로, 본원 명세서 내에 개시된 생물학적 시험 데이터는, 크레아틴 (Creatinine) 수치, 알부민 (Albumin) 수치, CRP (Creactive protein) 수치, WBC (white blood cell) 최소 수치, WBC 최고 수치, 혈소판 (Platelet) 수, 프로트롬빈 (Prothrombin) 시간 및 ALP (Alkaline phosphatase) 수치일 수 있다.
- [0069] 바람직하게 생물학적 시험 데이터는, ALP 수치, 혈소판 수치일 수 있으나, 이에 제한되는 것은 아니다.
- [0070] 본 명세서에서 사용되는 용어, "생체 신호 데이터"는 활력 징후, 생명 징후로, 개체의 상태와 연관된 데이터를



의미할 수 있다. 이러한 생체 신호 데이터는, 개체의 균혈증 발병 위험도와 연관될 수 있다.

- [0071] 이때, 생체 신호 데이터는 생체 신호 측정 장비로부터 측정된 개체에 대한 체온, 맥박, 산소 포화도, SBP (systolic blood pressure), DBP (diastolic blood pressure), 평균 혈압 및 호흡수일 수 있다. 바람직하게, 생체 신호 데이터는, SBP, 최고 체온, 최저 체온, 심박수 및 호흡수 중 적어도 하나일 수 있다. 그러나, 이에 제한되지 않고 생체 신호 데이터는, 개체의 건강 상태와 연관된 다양한 측정 데이터를 포함할 수 있다.
- [0072] 본 명세서에서 사용되는 용어, "임상적 특징 데이터"는 개체의 특징, 예를 들어 성별, 연령과 함께 약물 투여 여부, 의료적 처치 여부에 대한 데이터를 포괄하여 의미할 수 있다.
- [0073] 한편, 본원 명세서 내 개시된 임상적 특징 데이터는, 성별, 연령, 혈액 배양 기간, ICU (intensive care unit) 치료 여부, 중심 정맥 카테터 (Central venous catheter) 삽입 여부, 스테로이드 치료 여부 및 항생제 치료 여부 중 적어도 하나일 수 있다. 그러나, 임상적 특징 데이터는 이에 제한되지 않고, 개체에 투여되는 약물, 개체에 대한 의료적 처치에 대한 보다 다양한 데이터를 포함할 수 있다.
- [0074] 예를 들어, 임상적 특징 데이터는, 카테터 (Catheter), 폴리 (Foley), 기계환기 (Mechanical ventilator), 리스트레인트 (Restraint), 배액 기기 (Drainage) 중 적어도 하나의 의료적 처치 여부에 대한 데이터, 나아가 울트라셋 (Ultracet), 미다졸람 (Midazolam), 울티바 (ultiva), 포폴 (pofol), 아티반 (ativan), 펜타닐 (fentanyl), 프리세덱스 (precedex), IR 코돈 (IR codon), 타진 20/10 (TARGIN 20/10), 페리돌 (peridol), 리스페달 (risperdal), 자이프렉사 (zyprexa), 세로켈 (seroquel), 아빌리파이 (abilify), 페티딘 (pethidine), 듀로제식 패치 (durogesic patch), 모르핀 (morphine) 및 마이폴 (mypol) 의 투여 데이터를 더 포함할 수도 있다.
- [0075] 본 명세서에서 사용되는 용어, "균혈증 예측 모델"은 생물학적 시험 데이터, 생체 신호 데이터 및 임상적 특징 데이터 중 적어도 하나의 임상 데이터를 기초로 균혈증 발병의 위험도를 예측하도록 학습된 모델일 수 있다.
- [0076] 예를 들어, 균혈증 예측 모델은, 균혈증이 발병하거나 발병하지 않은 표본 개체로부터, 미리 결정된 시간마다 획득한 임상 데이터를 기초로 균혈증 발병의 위험도를 예측하도록 학습된 모델일 수 있다.
- [0077] 이때, 균혈증 예측 모델은, 생물학적 시험 데이터, 생체 신호 데이터 및 임상적 특징 데이터의 모든 임상 데이터를 기초로 균혈증 발병 위험도를 예측하도록 학습된 모델일 수 있다. 그러나, 학습에 이용되는 데이터는, 이에 제한되지 않으며 보다 다양한 임상 데이터들의 조합이 상기 예측 모델의 학습에 이용될 수 있다.
- [0078] 한편, 본 발명의 균혈증 예측 모델은, MLP (Multi-Layer Perceptron), SVM (support vector machine), RF (random forest), 및 MLP, SVM, RF의 조합으로 구성된 앙상블 모델, 선행 학습된 (pre-trained) 모델, Xgboost 모델 중 적어도 하나의 모델에 기초할 수 있다.
- [0079] 바람직하게, 본 발명의 균혈증 예측 모델은, 하나인 MLP (Multi-Layer Perceptron) 에 기초한 예측 모델일 수 있다. 이때, MLP 모델이 기초한 본 발명의 균혈증 예측 모델은, 128 노드 (nodes) 또는 256 노드를 갖는 히든 레이어 (hidden layer) 를 포함할 수 있다.
- [0080] 예를 들어, 본 발명의 균혈증 예측 모델은, 생물학적 시험 데이터, 생체 신호 데이터 및 임상적 특징 데이터가 입력되는 입력 레이어 및 균혈증 또는 균혈증이 아닌 정상을 예측하는 출력 레이어와 이들 레이어 사이에 128 노드 또는 256 노드를 갖는 히든 레이어가 존재하는 다층 레이어 구조의 예측 모델일 수 있다.
- [0081] 보다 구체적으로, 본 발명의 균혈증 예측 모델은 학습에 있어서 매개 변수를 갱신하는 최적화 함수로서 'rmsprop'를 이용하도록 구성될 수 있고, 다양한 임상 데이터들의 입력 값이 출력값에 전달되는 강도를 결정하는 함수로서 'relu' 함수를 이용하도록 구성될 수 있다.
- [0082] 그러나, 본 발명의 균혈증 예측 모델의 구성은 이에 제한되는 것이 아니다.
- [0083] 본 발명의 다른 실시예에 따르면, 본 발명의 균혈증 예측 모델은, 128 노드의 단일 노드를 갖는 MLP 모델일 수도 있다. 나아가, 본 발명의 균혈증 예측 모델은, 128 노드를 갖는 히든 레이어를 갖는 2 층의 히든 레이어를 갖는 MLP 모델일 수도 있다.
- [0084] 본 발명의 또 다른 실시예에 따르면, 본 발명의 균혈증 예측 모델은, gbtree 레이어를 갖는 Xgboost 모델일 수도 있다.
- [0085] 그러나, 본 발명의 균혈증 예측 모델의 종류는 전술한 것에 제한되는 것이 아니다. 예를 들어, 상기 예측 모델

은, 보다 다양한 ANN (Artificial Neural Network) 모델, 또는 DNN (Deep Neural Network), CNN (Convolutional Neural Network), RNN (Recurrent Neural Network), DCNN (Deep Convolutional Neural Network), RBM (Restricted Boltzmann Machine), DBN (Deep Belief Network), SSD (Single Shot Detector) 모델, U-net을 기반으로 하는 예측 모델일 수도 있다.

- [0086] 한편, 본 발명의 균혈증 예측 모델의 학습에 있어서, 균혈증이 아닌 정상을 예측하는 것에 영향력이 큰 임상 데이터들이 결정되는 평가가 수행될 수 있다. 이때, 영향력 평가는 OOS (one-out search), LRP (Layer-wise Relevance Propagation), GI (gradient input), IG (integrated gradients), 및 SM (saliency map) 중 적어도 하나의 알고리즘에 기초하여 수행될 수 있다.
- [0087] 예를 들어, 상기 적어도 하나의 알고리즘에 의해 생물학적 시험 데이터, 생체 신호 데이터 및 임상적 특징 데이터의 다양한 임상 데이터 중 균혈증임을 예측하는 것에 영향력이 큰 임상 데이터 또는, 균혈증이 아닌 정상을 예측하는 것에 영향력이 큰 임상 데이터가 결정될 수 있다.
- [0088] 이러한 영향력이 큰 임상 데이터들이 상기 예측 모델의 학습에 입력 데이터로서 적용될 수 있음에 따라, 예측 모델은 다른 모델들 보다 균혈증 예측 능력이 향상된 모델일 수 있다.
- [0089] 이하에서는 도 1a 및 1b를 참조하여, 본 발명의 일 실시예에 따른 균혈증 발병 위험도의 예측용 디바이스 및 이를 이용한 균혈증 발병 위험도 예측 시스템에 관하여 구체적으로 설명한다.
- [0090] 도 1a는 본 발명의 일 실시예에 따른 균혈증 발병 위험도의 예측 방법 및 디바이스에 기초한 균혈증 발병 위험도 예측 시스템을 예시적으로 도시한 것이다. 도 1b는 본 발명의 일 실시예에 따른 균혈증 발병 위험도의 예측용 디바이스의 구성을 예시적으로 도시한 것이다.
- [0091] 도 1a를 참조하면, 본 발명의 일 실시예에 따른 균혈증 발병 위험도 예측 시스템 (1000) 은, 균혈증 발병 위험도 예측용 디바이스 (100), 개체 (200) 에 대하여 획득된, 생물학적 시험 데이터 (310), 생체 신호 데이터 (320), 임상적 특징 데이터 (330) 를 포함하는 임상 데이터 (300), 의료적 처치 기구 (400) 및 의료진 디바이스 (500) 로 구성되어 있다.
- [0092] 이때, 임상 데이터 (300) 는 개체 (200) 로부터, 임의의 시점에서 평가된 데이터일 수 있다. 그러나, 이에 제한되는 것은 아니다.
- [0093] 보다 구체적으로, 균혈증 발병 위험도 예측 시스템 (1000) 에서 균혈증 발병 위험도 예측용 디바이스 (100) 는, 개체 (200) 에 대하여 측정되거나 평가된 다양한 임상 데이터 (300) 를 수신하고, 이를 기초로 균혈증 발병 위험도를 예측하도록 구성될 수 있다.
- [0094] 이때, 의료적 처치 기구 (400) 는 개체 (200) 에 온도, 맥박, 산소 포화도, 수축기 혈압, 이완기 혈압, 평균 혈압 및 호흡수로 이루어진 그룹 중 적어도 하나의 생체 신호 데이터 (320) 를 제공할 수 있는 생체 신호 측정 디바이스일 수 있다.
- [0095] 한편, 균혈증 발병 위험도 예측 시스템 (1000) 에서 생물학적 시험 데이터 (310), 나아가 임상적 특징 데이터 (330) 는 EMR (Electronic Medical Record) 시스템과 같은 외부 시스템으로부터 획득될 수도 있다.
- [0096] 보다 구체적으로, 도 1b를 참조하면, 균혈증 발병 위험도 예측용 디바이스 (100) 는 수신부 (110), 입력부 (120), 출력부 (130), 저장부 (140) 및 프로세서 (150) 를 포함한다.
- [0097] 구체적으로 수신부 (110) 는 개체 (200) 에 대한 생물학적 시험 데이터 (310), 생체 신호 데이터 (320), 임상적 특징 데이터 (330) 를 포함하는 임상 데이터 (300) 를 수신하도록 구성될 수 있다.
- [0098] 본 발명의 특징에 따르면, 수신부 (110) 는 의료진 디바이스 (500) 에 대하여, 후술될 프로세서 (150) 에 의해 결정된 개체 (200) 에 대하여 예측된 결과를, 송신하도록 더 구성될 수 있다.
- [0099] 입력부 (120) 는 키보드, 마우스, 터치 스크린 패널 등 제한되지 않는다. 입력부 (120) 는 균혈증 발병 위험도 예측용 디바이스 (100) 를 설정하고, 균혈증 발병 위험도 예측용 디바이스 (100) 의 동작을 지시할 수 있다.
- [0100] 한편, 출력부 (130) 는 수신부 (110) 에 의해 수신된 다양한 임상 데이터 (300) 를 표시할 수 있다. 나아가, 출력부 (130) 는 프로세서 (150) 에 의해 예측된 균혈증 진단과 연관된 정보를 디스플레이 적으로 표시할 수 있다. 나아가, 출력부 (130) 는 프로세서 (150) 에 의해 균혈증 발병 위험도가 높은 것으로 결정될 경우, 알람 소리를 출력하도록 더 구성될 수 있다.

- [0101] 저장부 (140) 는 수신부 (110) 를 통해 수신한 개체 (200) 에 대한 다양한 임상 데이터 (300) 를 저장하고, 입력부 (120) 를 통해 설정된 균혈증 발병 위험도 예측용 디바이스 (100) 의 지시를 저장하도록 구성될 수 있다. 나아가, 저장부 (140) 는 후술될 프로세서 (150) 에 의해 생성된 개체 (200) 에 균혈증 예측 정보를 저장하도록 구성된다. 그러나, 전술한 것에 제한되지 않고 저장부 (140) 는 균혈증 발병 위험도 예측을 위해 프로세서 (150) 에 의해 결정된 다양한 정보들을 저장할 수 있다.
- [0102] 프로세서 (150) 는 균혈증 발병 위험도 예측용 디바이스 (110) 의 정확한 예측 결과를 제공하기 위한 구성 요소일 수 있다. 이때, 정확한 균혈증 발병 위험도 예측을 위해 프로세서 (150) 는 다양한 임상 데이터 (300) 를 기초로 균혈증 발병 위험도를 예측하도록 구성된 예측 모델에 기초할 수 있다.
- [0103] 한편, 프로세서 (150) 는, 개체 (200) 가 균혈증 예측 모델에 의해 균혈증 발병 고위험군으로 예측된 경우 의료진 디바이스 (500) 에 알림을 제공하도록 구성될 수도 있다. 이에, 의료진은 개체 (200) 에 대한 빠른 조치를 취할 수 있다.
- [0104] 이와 같이 본 발명의 균혈증 발병 위험도 예측 시스템 (1000) 은, 개체 (200) 에 대하여 균혈증의 조기 진단 및 균혈증의 조기 치료를 제공할 수 있음에 따라, 다양한 의료 시스템에 적용될 수 있다.
- [0105] 이하에서는 도 2를 참조하여, 본 발명의 일 실시예에 따른 균혈증 발병 위험도의 예측 방법을 구체적으로 설명한다. 도 2는, 본 발명의 일 실시예에 따른 균혈증 발병 위험도의 예측 방법의 절차를 예시적으로 도시한 것이다.
- [0106] 도 2를 참조하면, 본 발명의 일 실시예에 따른 균혈증 발병 위험도의 예측 방법은, 먼저 개체로부터 생물학적 시험 데이터를 수신하고 (S210), 생물학적 시험 데이터를 기초로 균혈증 발병 위험도를 예측하도록 구성된 균혈증 예측 모델을 이용하여, 개체에 대한 균혈증 발병 위험도를 예측한다 (S220). 최종적으로 개체에 대하여 예측된 균혈증 발병 위험도를 제공한다 (S230).
- [0107] 보다 구체적으로, 데이터를 수신하는 단계 (S210) 에서, 크레아틴 (Creatinine) 수치, 알부민 (Albumin) 수치, CRP (Creactive protein) 수치, WBC (white blood cell) 최소 수치, WBC 최고 수치, 혈소판 (Platelet) 수, 프로트롬빈 (Prothrombin) 시간 중 적어도 하나 및 ALP (Alkaline phosphatase) 수치의 생물학적 시험 데이터가 수신될 수 있다.
- [0108] 본 발명의 특징에 따르면, 데이터를 수신하는 단계 (S210) 에서, SBP (systolic blood pressure), 최고 체온, 최저 체온, 심박수 및 호흡수 중 적어도 하나의 생체 신호 데이터가 더 수신될 수 있다.
- [0109] 본 발명의 다른 특징에 따르면, 데이터를 수신하는 단계 (S210) 에서, 성별, 연령, 혈액 배양 기간, ICU (intensive care unit) 치료 여부, 중심 정맥 카테터 (Central venous catheter) 삽입 여부, 스테로이드 치료 여부 및 항생제 치료 여부 중 적어도 하나의 임상적 특징 데이터가 더 수신될 수 있다.
- [0110] 본 발명의 특징에 따르면, 데이터를 수신하는 단계 (S210) 에서 수신된 생체 신호 데이터는, ALP 수치, 혈소판 수일 수 있으나, 이에 제한되는 것은 아니다.
- [0111] 다음으로, 균혈증 발병 위험도를 예측하는 단계 (S220) 에서는, 균혈증 발병 위험도를 예측하도록 구성된 균혈증 예측 모델에 의해 개체가 균혈증 또는 균혈증이 아닌 정상으로 결정될 수 있다.
- [0112] 이때, 균혈증 예측 모델은, 데이터를 수신하는 단계 (S210) 에서 수신된, 생물학적 시험 데이터, 생체 신호 데이터, 및 임상적 특징 데이터 중 적어도 하나의 임상 데이터를 기초로 균혈증 발병의 위험도를 예측하도록 학습된 모델일 수 있다.
- [0113] 예를 들어, 균혈증 예측 모델은 균혈증 발병 표본 개체 및 정상의 표본 개체에 대한 생물학적 시험 데이터, 생체 신호 데이터, 및 임상적 평가 데이터 중 적어도 하나의 학습용 데이터를 수신하고, 학습용 데이터를 기초로, 균혈증 또는 정상을 예측 또는 분류하는 단계를 통해 학습된 모델일 수 있다.
- [0114] 나아가, 균혈증 예측 모델은, 학습용 데이터에 대하여 균혈증과의 관련도 (relevance) 점수를 산출하고, 관련도 점수를 기초로, 미리 결정된 순위 내에 있는 균혈증 관련 학습용 데이터를 결정하는 단계를 통해 결정된 균혈증 관련 학습용 데이터를 기초로, 균혈증 발병 위험도를 예측하도록 구성된 모델일 수 있다.
- [0115] 본 발명의 다른 특징에 따르면, 본 발명의 균혈증 예측 모델은, MLP (multi-layer perceptron), SVM (support vector machine), RF (random forest), 및 상기 MLP, 상기 SVM, 상기 RF의 조합으로 구성된 앙상블 모델 중 적

어도 하나의 모델일 수 있다.

- [0116] 본 발명의 또 다른 특징에 따르면, 균혈증 예측 모델은, 128 노드 (nodes) 또는 256 노드를 갖는 히든 레이어 (hidden layer) 로 구성된 MLP 모델일 수 있으나, 이에 제한되는 것은 아니다.
- [0117] 마지막으로, 균혈증 발병 위험도를 제공하는 단계 (S230) 에서는, 전술한 균혈증 발병 위험도를 예측하는 단계 (S220) 에서 예측된 균혈증 진단에 관한 정보를 제공할 수 있다.
- [0118] 본 발명의 다른 특징에 따르면, 균혈증 발병 위험도를 제공하는 단계 (S230) 에서는, 균혈증 발병 위험도를 예측하는 단계 (S220) 에서 균혈증 예측 모델에 의해 개체에 대한 균혈증 발병 위험도가 예측될 경우, 개체에 대한 균혈증 발병의 위험 알람이 제공될 수 있다.
- [0119] 한편, 균혈증 발병 위험도를 제공하는 단계 (S230) 에서는, 전술한 것 이외에 보다 다양한 정보가 제공될 수 있다. 예를 들어, 균혈증 발병 위험도를 제공하는 단계 (S230) 에서는, 균혈증 교정에 효과적일 수 있는 추가적 처치 정보를 더 제공할 수 있다.
- [0120] 이상의 절차에 따라, 본 발명의 일 실시예에 따른 균혈증 발병 위험도 예측 방법은, 실시간으로 개체에 대한 균혈증 발병 여부를 결정하거나, 균혈증 발병의 위험도를 미리 예측하여, 균혈증 발병 위험도에 대한 정보, 나아가 균혈증 고위험군을 알리는 알람을 제공할 수 있다. 이에, 의료진은 개체에 대한 균혈증의 조기 진단이 가능할 수 있다. 나아가, 의료진은 균혈증 고위험군에 대한 빠른 조치를 취할 수 있다.
- [0121] 이하에서는, 도 3a 내지 3d를 참조하여, 본 발명의 다양한 실시예에 이용되는 균혈증 예측 모델에 대하여 구체적으로 설명한다.
- [0122] 도 3a 및 3b는 본 발명의 다양한 실시예에 이용되는 균혈증 예측 모델의 학습 데이터를 예시적으로 도시한 것이다.
- [0123] 도 3a를 참조하면, 본 발명의 다양한 실시예에 이용되는 균혈증 예측 모델의 학습을 위해 강남세브란스 병원으로부터 획득한 13402 명의 환자로부터 획득한 22332 개의 혈액 배양 에피소드 (episode) 를 수집하였고, 이들 중 1260 개의 균혈증 개체에 대한 에피소드 및 1260 개의 정상 개체에 대한 에피소드를 학습용 데이터로 설정하였다. 이에, 본 발명의 균혈증 예측 모델은, 1 : 1 비율의 양성 결과 및 음성 결과 데이터를 기초로 균혈증을 예측하도록 학습되었다.
- [0124] 도 3b를 참조하면, 균혈증 발병 개체 및 정상 개체의 표본 개체로부터 획득된 학습용 데이터는, 생체 신호 데이터, 생물학적 시험 데이터, 및 임상적 특징 데이터로 구성된다. 이들 데이터 각각의 값은, 균혈증 발병 개체 및 정상 개체에 대하여 통계적으로 유의한 수준으로 차이 나는 것으로 나타난다.
- [0125] 보다 구체적으로, 학습용 생체 신호 데이터는 SBP, 최고 체온, 최저 체온, 심박수 및 호흡수로 구성될 수 있다. 학습용 생물학적 시험 데이터는, 크레아틴 수치, 알부민 수치, CRP 수치, WBC 최소 수치, WBC 최고 수치, 혈소판 수, 프로트롬빈 시간 및 ALP 수치로 구성될 수 있다. 나아가, 임상적 특징 데이터는, 혈액 배양 기간, ICU 치료 여부, 중심 정맥 카테터 여부, 스테로이드 치료 여부, 항생제 치료 여부, 연령 및 성별로 구성될 수 있다.
- [0126] 이때, 임상적 특징 데이터 중, ICU 치료 여부 및 중심 정맥 카테터 처치 여부는 혈액 배양 당시에 평가된 데이터일 수 있고, 항생제 치료 여부는 혈액 배양 3주 이전에 평가된 데이터일 수 있으며, 스테로이드 치료 여부는 혈액 배양 2주 이전에 평가된 데이터일 수 있다. 이때, 생체 신호 데이터 및 생물학적 시험 데이터는 혈액 배양 72 시간 이전에 평가된 데이터일 수 있다.
- [0127] 본 발명의 균혈증 예측 모델은 상기와 같은 학습용 데이터를 기초로 표본 개체에 대한 균혈증의 발병 여부를 결정하도록 학습될 수 있다. 도 3c는 본 발명의 다양한 실시예에 이용되는 균혈증 예측 모델의 구성을 예시적으로 도시한 것이다.
- [0128] 도 3c를 참조하면, 본 발명의 균혈증 예측 모델은 다층 인공 신경망인 MLP 알고리즘에 기초한 예측 모델일 수 있다.
- [0129] 보다 구체적으로, 본 발명의 균혈증 예측 모델은, ALP 수치 또는 혈소판 수치와 같은 생물학적 시험 데이터와 최고 체온과 같은 생체 신호 데이터, 임상적 특징 데이터가 입력되는 입력 레이어 및 균혈증 또는 균혈증이 아닌 정상을 예측하는 출력 레이어와 이들 레이어 사이에 1 개의 히든 레이어가 존재하는 다층 구조의 예측 모델일 수 있다.



- [0130] 이때, 본 발명의 균열증 예측 모델은, 128 개 또는 256 개의 노드로 구성된 1 개의 히든 레이어를 포함하도록 구성될 수 있다. 나아가, 상기 예측 모델은, 균열증의 예측 학습에 있어서 예측의 오차를 최소화하는 가중치를 찾기 위한 파라미터일 수 있는 학습 비율 (learning rate) 값이 0.1로 설정될 수 있다. 또한, 예측의 오차를 최소화하며 학습 속도를 증가시키기 위한 파라미터 값인 모멘텀 (momentum) 값이 0.95로 설정될 수 있다. 또한, 본 발명의 균열증 예측 모델은, 학습에 있어서 매개 변수를 갱신하는 최적화 함수로서 'adadelta'를 이용하도록 구성될 수 있고, 다양한 임상 데이터들의 입력 값이 출력값에 전달되는 강도를 결정하는 함수로서 'tanh' 함수를 이용하도록 구성될 수 있다.
- [0131] 그러나, 본 발명의 균열증 예측 모델의 종류는 이에 제한되는 것이 아니다. 예를 들어, 상기 예측 모델은, SVM (support vector machine), RF (random forest), 및 MLP, SVM, RF의 조합으로 구성된 앙상블 모델 중 적어도 하나의 모델일 수 있고, DNN, CNN, RNN, DCNN, RBM, DBN, SSD 모델 또는 U-net을 기반으로 하는 예측 모델일 수도 있다.
- [0132] 한편, 본 발명의 균열증 예측 모델의 학습에 있어서, 균열증이 아닌 정상을 예측하는 것에 영향력이 큰 임상 데이터들이 결정되는 평가가 수행될 수 있다. 이때, 영향력 평가는 OOS (one-out search), LRP (Layer-wise Relevance Propagation), GI (gradient input), IG (integrated gradients), 및 SM (saliency map) 중 적어도 하나의 알고리즘에 기초하여 수행될 수 있다.
- [0133] 도 3d는 본 발명의 다양한 실시예에 이용되는 균열증 예측 모델의 평가 방법을 예시적으로 도시한 것이다.
- [0134] 이때, LRP 알고리즘에 기초한 임상 데이터의 평가를 예로 들어 설명하나, 평가 알고리즘은 전술한 바와 같이 보다 다양한 알고리즘에 기초하여 수행될 수 있다.
- [0135] 도 3d를 참조하면, 다양한 임상 데이터에 기초하여 균열증 예측 모델에 의해 균열증 예측 결과가 출력되면, LRP 알고리즘에 기초한 평가가 수행될 수 있다.
- [0136] 보다 구체적으로, 본 평가에서는, LRP 알고리즘에 의해 ALP 수치, 혈소판 수와 같은 생물학적 시험 데이터, 최고 체온과 같은 생체 신호 데이터, 나아가 임상적 특징 데이터의 다양한 임상 데이터의 입력값과 균열증 또는 균열증이 아닌 정상으로의 출력값의 관련도 (relevance) 가 추정될 수 있다.
- [0137] 보다 구체적으로, 다양한 임상 데이터의 입력값과 균열증 또는 균열증이 아닌 정상으로의 출력값의 관련도는 하기 [수학식 1]로 산출될 수 있다.
- [0138] [수학식 1]
- $$r_i^{(L)} = \begin{cases} S_i(x) & \text{if unit } i \text{ is the target unit of interest} \\ 0 & \text{otherwise} \end{cases}$$
- $$r_i^{(l)} = \sum_j \frac{z_{ji}}{\sum_{j'} (z_{ji'} + b_j) + \epsilon \cdot \text{sign}(\sum_{j'} (z_{ji'} + b_j))} r_j^{(l+1)}$$
- [0139]
- [0140] 여기서,  $L$ 은, 출력 레이어 값을 의미하고,  $Z_{ji}$ 는  $l+1$ 번째 레이어와  $l$ 번째 레이어의 가중치, 및  $l$ 번째 레이어 입력 값의 곱을 의미할 수 있다 ( $w_{ji}^{(l+1,l)} x_i^{(l)}$ ). 나아가,  $b_j$ 는  $j$ 번째 노드의 편향값을 의미할 수 있다.
- [0141] 산출된 관련도 점수에 의해 다양한 임상 데이터 중, 패혈증과 상이한 균열증임을 예측하는 것에 관련도가 큰 임상 데이터 또는, 균열증이 아닌 정상을 예측하는 것에 영향력이 큰 임상 데이터가 결정될 수 있다.
- [0142] 이러한 영향력이 큰 임상 데이터들이 상기 예측 모델의 학습에 입력 데이터로서 적용될 수 있음에 따라, 균열증의 예측에 있어서 본 발명의 예측 모델은 다른 모델을 보다 균열증 예측 능력이 우수할 수 있다.
- [0143] **실시예 1: 본 발명의 다양한 실시예에 따른 균열증 예측 모델의 제1 평가**
- [0144] 이하에서는 4a 내지 4f를 참조하여 본 발명의 다양한 실시예에서 이용되는 균열증 예측 모델의 제1 평가 결과에 대하여 구체적으로 설명한다. 도 4a 및 4b는 본 발명의 다양한 실시예에 이용되는 균열증 예측 모델에 대한 제1 평가 결과를 도시한 것이다.
- [0145] 먼저, 도 4a를 참조하면, 생물학적 시험 데이터, 생체 신호 데이터 및 데이터를 기초로 균열증 또는 균열증이 아닌 정상을 예측하도록 구성된 균열증 예측 모델의 평가 결과가 도시된다.
- [0146] 이때, 본 발명의 균열증 예측 모델은, 128 개의 노드를 갖는 히든 레이어로 구성된 MLP 모델 (본 발명의 균열증

예측 모델 (A)), 256 개의 노드를 갖는 히든 레이어로 구성된 MLP 모델 (본 발명의 균혈증 예측 모델 (B)), 랜덤 포레스트의 분류 모델 (본 발명의 균혈증 예측 모델 (C)), SVM의 분류 모델 (본 발명의 균혈증 예측 모델 (D)) 및 128 개의 노드를 갖는 히든 레이어로 구성된 MLP 모델, 256 개의 노드를 갖는 히든 레이어로 구성된 MLP 모델 및 랜덤 포레스트 분류 모델이 결합된 앙상블 모델 (본 발명의 균혈증 예측 모델 (E)) 로 구성된다. 나아가, 각 변수가 계층화되어 섹션으로 나눈 후 균혈증에 대한 위험을 분석하도록 구성된 종래의 예측 모델 Naïve Bayesian이 본 평가에 함께 이용되었다.

[0147] 도 4a를 참조하면, MLP에 기초한, 본 발명의 균혈증 예측 모델 (A) 및 본 발명의 균혈증 예측 모델 (B) 의 민감도가 0.81로 나머지 모델에 비하여 가장 높은 것으로 나타난다. 나아가, 본 발명의 균혈증 예측 모델 (C) 의 특이도가 0.655로 나머지 모델에 비하여 가장 높은 것으로 나타난다.

[0148] 특히, 도 4b를 함께 참조하면 본 발명의 균혈증 예측 모델 (A) 내지 (E) 의 진단 능력을 의미할 수 있는 AUC값은 평균 약 0.729로, 종래의 예측 모델인 Naïve Bayesian에 비하여 높은 것으로 나타난다. 이때, AUC 값은 진단 능력과 연관이 있을 수 있다. 따라서, 본 발명의 다양한 실시예에 따른 균혈증 예측 모델은, 균혈증 진단 시스템에 적용될 경우, 개체에 대하여 균혈증의 발병 위험도를 높은 정확도로 미리 예측할 수 있다.

[0149] 이상의 결과에 따르면, 본 발명의 다양한 실시예에 따른 균혈증 예측 모델이 종래의 예측 모델보다 균혈증 발병의 위험도를 높은 신뢰도로 예측하는 것으로 나타난다. 특히, MLP 기반의 균혈증 예측 모델의 경우, 다른 모델들 보다 높은 민감도를 유지하면서 높은 AUC 값을 갖는 것으로 나타난다. 이에, 본 발명의 균혈증 예측 모델은, MLP 기반의 예측 모델일 수 있으나 이에 제한되는 것은 아니다.

[0150] 도 4c 및 4d는 본 발명의 다양한 실시예에 이용되는 균혈증 예측 모델에 대한, 균혈증 예측의 영향력에 따른 임상 데이터들을 도시한 것이다.

[0151] 도 4c를 참조하면, 본 발명의 다양한 실시예에서 이용되는 임상 데이터들 (생물학적 시험 데이터, 생체 신호 데이터 및 임상적 특징 데이터) 의 영향력 (관련도) 평가 결과가 도시된다. 본 평가에서는, 'One-out search' 및 'Gini Importance'에 기초하여 균혈증의 예측에 영향력이 높은 변수들의 순위를 평가하였다. 보다 구체적으로, MLP 기반의 예측 모델에 적용될 수 있는 One-out search의 순위 평가 결과에 따르면, 생물학적 시험 데이터 중 ALP 수치 및 혈소판 수와, 생체 신호 데이터 중 최고 체온이 균혈증의 발병 위험도를 예측하는 것에 있어서 높은 영향력이 있는 것으로 나타난다. 한편, 랜덤 포레스트 기반의 예측 모델에 이용될 수 있는 Gini Importance의 순위 평가 결과에 따르면, One-out search의 순위 평가 결과와 다소 상이하지만, One-out search에 의해 영향력이 높다고 평가된 ALP 수치, 혈소판 수 및 최고 체온이 균혈증을 분류 하는 것에 높은 영향력을 갖는 것으로 나타난다.

[0152] 도 4d를 더욱 참조하면, 그레디언트 기반의 방법의 LRP (Layer-wise Relevance Propagation), GI (gradient input), IG (integrated gradients), 및 SM (saliency map) 에 기초하여 균혈증의 예측에 영향력이 높은 변수들의 순위를 평가한 결과가 도시된다. 보다 구체적으로, LRP, GI 및 IG의 순위 평가 결과에 따르면, 전술한 One-out search의 순위 평가 결과와 유사하게 ALP 수치가 균혈증의 발병 위험도를 예측하는 것에 높은 영향력을 갖는 것으로 나타난다. 또한, 혈소판 수가 균혈증의 발병 위험도를 예측하는 것에 높은 영향력을 갖는 것으로 나타난다. SM에 따른 순위 평가 결과에서도, 혈소판 수 및 ALP 수치가 균혈증의 발병 위험도를 예측하는 것에 높은 영향력을 갖는 것으로 나타난다.

[0153] 이러한 결과는, 개체에 대한 임상 데이터를 기초로 균혈증을 예측 하는 것에 있어서, 생물학적 시험 데이터 중 ALP 수치 및 혈소판 수, 나아가 생체 신호 데이터 중 최고 체온이 균혈증 예측에 있어서 큰 영향력을 제공한다는 것을 의미할 수 있다. 즉, 본 발명의 균혈증 예측 모델은, 중 ALP 수치 및 혈소판 수, 나아가 최고 체온을 기초로 균혈증 발병 여부를 예측하도록 학습될 수 있어, 우수한 진단 능력으로 개체에 대한 균혈증 발병의 위험도를 예측할 수 있다.

[0154] 도 4e 및 4f는 본 발명의 다양한 실시예에 이용되는 균혈증 예측 모델에 대한, 균혈증 예측의 영향력에 따른 임상 데이터들의 변화에 따른 성능 변화를 도시한 것이다.

[0155] 도 4e의 (a)를 참조하면, MLP의 ANN 머신 러닝 알고리즘에 대하여 변수의 평가를 위해 LRP, SM, OOS (one-out search), 및 교란-기반 방법의 Occlusion을 적용한 후, 영향력이 높다고 평가된 변수들을 영향력 순으로 누적하여 제외했을 때의 AUC값이 도시된다. 나아가, RF 기반의 예측 모델에 대하여 변수의 평가를 위해 Gini Importance 방법으로 영향력이 높다고 평가된 변수들을 영향력 순으로 누적하여 제외했을 때의 AUC값이 함께 도

시된다.

- [0156] 보다 구체적으로, 균혈증을 예측하는 것에 있어서 영향력이 높은 것으로 평가된 상위 변수들 (예를 들어, ALP 수치, 혈소판 수, 최고 체온 등) 이 제외됨에 따라, 균혈증 예측에 대한 AUC값이 감소하는 것으로 나타난다. 이는 영향력이 높은 임상 데이터들의 제외에 따른, 균혈증 예측의 성능 저하를 의미할 수 있다.
- [0157] 도 4e의 (b)를 참조하면, MLP의 ANN 머신 러닝 알고리즘에 대하여 변수의 평가를 위해 LRP, SM, OOS (one-out search), 및 교란-기반 방법의 Occlusion을 적용한 후, 영향력이 높다고 평가된 변수들을 영향력 순으로 누적하여 추가했을 때의 AUC값이 도시된다. 나아가, RF 기반의 예측 모델에 대하여 변수의 평가를 위해 Gini Importance 방법으로 영향력이 높다고 평가된 변수들을 영향력 순으로 누적하여 추가했을 때의 AUC값이 함께 도시된다.
- [0158] 보다 구체적으로, 균혈증을 예측하는 것에 있어서 영향력이 높은 것으로 평가된 상위 변수들 (예를 들어, ALP 수치, 혈소판 수, 최고 체온 등) 이 추가됨에 따라, 균혈증 예측에 대한 AUC값이 증가하는 것으로 나타난다. 이때, 상위 10 개의 변수를 추가한 이후 AUC 값의 변화는 미미한 것으로 나타난다. 이러한 결과는, 상위 10 개의 임상 데이터들이 균혈증을 예측하는 것에 있어서 영향력이 높은 것을 의미할 수 있다. 한편, OOS의 경우, 다른 평가 모델들보다 변수 추가에 따른 현저한 AUC 값의 증가가 나타난다. 이에, 본 발명의 다양한 실시예에 따른 균혈증 발병 위험도 예측 방법에 있어서, 다양한 임상적 변수들, 즉 생물학적 시험 데이터, 생체 신호 데이터 및 임상적 특징 데이터의 평가에 OOS 모델이 적용될 수 있으나, 이에 제한되는 것은 아니다.
- [0159] 도 4f의 (a)를 참조하면, RF 및 SVM의 비-ANN 머신 러닝 알고리즘에 대하여 변수의 평가를 위해 Gini Importance 방법, 또는 OOS를 적용한 후, 영향력이 높다고 평가된 변수들을 영향력 순으로 누적하여 제외했을 때의 AUC값이 도시된다.
- [0160] 보다 구체적으로, 균혈증을 예측하는 것에 있어서 영향력이 높은 것으로 평가된 상위 변수들 (예를 들어, ALP 수치, 혈소판 수, 최고 체온 등) 이 제외됨에 따라, 균혈증 예측에 대한 AUC값이 감소하는 것으로 나타난다. 이는 영향력이 높은 임상 데이터들의 제외에 따른, 균혈증 예측의 성능 저하를 의미할 수 있다.
- [0161] 도 4f의 (b)를 참조하면, RF 및 SVM의 비-ANN 머신 러닝 알고리즘에 대하여 변수의 평가를 위해 Gini Importance 방법, 또는 OOS를 적용한 후, 영향력이 높다고 평가된 변수들을 영향력 순으로 누적하여 추가했을 때의 AUC값이 도시된다.
- [0162] 보다 구체적으로, 균혈증을 예측하는 것에 있어서 영향력이 높은 것으로 평가된 상위 변수들 (예를 들어, ALP 수치, 혈소판 수, 최고 체온 등) 이 추가됨에 따라, 균혈증 예측에 대한 AUC값이 증가하는 것으로 나타난다. 이때, 상위 10 개의 변수를 추가한 이후 AUC 값의 변화는 미미한 것으로 나타난다. 이러한 결과는, 상위 10 개의 임상 데이터들이 균혈증을 예측하는 것에 있어서 영향력이 높은 것을 의미할 수 있다. 한편, OOS의 경우, 다른 평가 모델들보다 변수 추가에 따른 현저한 AUC 값의 증가가 나타난다. 이에, 본 발명의 다양한 실시예에 따른 균혈증 발병 위험도 예측 방법에 있어서, 다양한 임상적 변수들, 즉 생물학적 시험 데이터, 생체 신호 데이터 및 임상적 특징 데이터의 평가에 OOS 모델이 적용될 수 있으나, 이에 제한되는 것은 아니다.
- [0163] 한편, 본 발명의 다양한 실시예에서 이용되는 균혈증 예측 모델은, 균혈증 예측에 있어서 영향력이 큰 임상 데이터를 학습 데이터로 이용하도록 구성될 수 있다. 이에, 균혈증 예측 모델은, 균혈증 예측에 있어서 다른 예측 모델보다 향상된 진단 능력으로 균혈증 예측 결과를 제공할 수 있다. 이상의 실시예 1의 결과로, 본 발명의 다양한 실시예에서 이용된 균혈증 예측 모델이 균혈증 발병 위험도를 높은 정확도로 예측하는 것을 확인할 수 있었다. 나아가, 높은 AUC 값을 갖는 것으로 나타난 결과에 따라 본 발명의 예측 모델은 우수한 진단 능력을 가지는 것으로 확인할 수 있었다.
- [0164] 이에, 본 발명은 균혈증에 대한 처치 시점을 앞당겨 치료에 대한 좋은 예후를 제공할 수 있다. 나아가, 본 발명은, 패혈증과 상이한 균혈증에 대하여 높은 신뢰도 및 정확도로 발병의 위험도를 예측할 수 있는 효과가 있다.
- [0165] 특히, 본 발명은 개체로부터 빠르게 획득 가능한 임상 데이터에 기초한 균혈증 발병 위험도 예측 시스템을 제공함으로써, 개체에 대하여 균혈증 발병의 조기 진단을 제공할 수 있다.
- [0166] 이에, 본 발명은 균혈증 발병 개체의 생존률 증가, 합병증 예방, 및 치료 비용의 감소 효과를 제공할 수 있다.
- [0167] 또한, 본 발명은, 개체에 대하여 예측 모델에 의해 균혈증이 예측될 경우, 알람을 제공하여 균혈증 발병 위험도를 알리도록 구성됨에 따라, 보호자 또는 의료진은 중환자와 같은 지속적인 모니터링이 요구되는 개체에 대하여

보다 용이하게 균혈증 발병 고위험군을 인지할 수 있다.

[0168] 실시예 2: 본 발명의 다양한 실시예에 따른 균혈증 예측 모델의 제2 평가

[0169] 이하에서는 5a 내지 5i를 참조하여 본 발명의 다양한 실시예에서 이용되는 균혈증 예측 모델의 제2 평가 결과에 대하여 구체적으로 설명한다. 도 5a 및 5b는 본 발명의 다양한 실시예에 이용되는 균혈증 예측 모델의 제2 평가를 위한 학습 데이터 및 평가 데이터를 도시한 것이다. 도 5c 내지 5i는 본 발명의 다양한 실시예에 이용되는 균혈증 예측 모델에 대한 제2 평가 결과를 도시한 것이다.

[0170] 이때, 도 5a를 참조하면, 본 평가에서 균혈증 예측을 위해, 임상적 특징 데이터로서 연령, 성별이 특징 파라미터로 설정 되었고, 생체 신호 데이터로서 SBP (systolic blood pressure), DBP (diastolic blood pressure), 최고 체온, 최저 체온, 심박수 및 호흡수가 특징 파라미터로 설정 되었다. 나아가, 생물학적 시험 데이터로서 알부민 수치, ALT (alanine aminotransferase) 수치, 트롬보플라스틴 (thromboplastin) 의 응고 시간, AST (aspartate aminotransferase) 수치, CRP (C-reactive protein) 수치, 적혈구 침강 속도, FERR (ferritin) 수치, HMG (hemoglobin) 수치, PTINR (Prothrombin Time International Normalized Ratio), PTPER (Prothrombin Time (%)), PTSEC (Prothrombin Time (sec)), 적혈구 수치, TBIL (Total bilirubin) 수치, TCO<sub>2</sub> (total carbon dioxide) 수치 및 백혈구 수치가 특징 파라미터로 설정되었다.

[0171] 이때, 생체 신호 데이터 및 생물학적 시험 데이터는, 3일 이내 측정된 최소값, 최대값, 또는 측정된 수치를 포함할 수 있다.

[0172] 보다 구체적으로, 도 5b를 참조하면, 본 발명의 다양한 실시예에 이용되는 균혈증 예측 모델의 제2 평가를 위해 이용된 데이터 세트 A, B 및 C의 구성이 도시된다.

[0173] A 데이터 세트는, 트레이닝 및 검증을 위해 균혈증 양성 및 균혈증 음성 데이터가 1:1의 비율을 갖도록 구성되고, 테스트를 위해 균혈증 양성 및 균혈증 음성 데이터가 약 1:10의 비율을 갖도록 구성된다. 이때, A 데이터 세트는, 강남 세브란스 병원 및 신촌 세브란스 병원에서 2007년 내지 2018년 사이에 샘플링된 균혈증 양성 및 균혈증 음성 데이터로서, 트레이닝 및 검증을 위해 2007년 내지 2015년도의 데이터가, 테스트를 위해 2016년 내지 2018년 데이터가 이용될 수 있다.

[0174] B 데이터 세트는, 트레이닝 및 검증을 위해 균혈증 양성 및 균혈증 음성 데이터가 1:1의 비율을 갖도록 구성되고, 테스트를 위해 균혈증 양성 및 균혈증 음성 데이터가 약 1:7의 비율을 갖도록 구성된다. 이때, C 데이터 세트는, 강남 세브란스 병원 및 신촌 세브란스 병원에서 2007년 내지 2018년 사이에 샘플링된 균혈증 양성 및 균혈증 음성 데이터를 연도별로 병합한 후 랜덤 분할된 데이터일 수 있다.

[0175] C 데이터 세트는, 트레이닝 및 검증을 위해 균혈증 양성 및 균혈증 음성 데이터가 1:1의 비율을 갖도록 구성되고, 테스트를 위해 균혈증 양성 및 균혈증 음성 데이터가 약 1:7의 비율을 갖도록 구성된다. 이때, C 데이터 세트는, 강남 세브란스 병원 및 신촌 세브란스 병원에서 2007년 내지 2018년 사이에 샘플링된 균혈증 양성 및 균혈증 음성 데이터를 연도별로 병합한 후 랜덤 분할하고, 동일 환자 샘플이 각 세트별로 중복되지 않도록 구분되고, 동일 환자의 양성/음성 샘플이 동시에 존재하는 경우 음성 샘플이 제거된 데이터일 수 있다.

[0176] 도 5c를 참조하면, 전술한 A 데이터에 기초하여 학습된 균혈증 예측 모델에 대한 평가 결과가 도시된다. 이때, 128 개의 노드, 64 개의 노드, 25 개의 노드 및 128 개의 노드로 구성된 4 층의 복수의 레이어로 구성된 균혈증 예측 모델에 대하여, 선행 학습여부에 따른 균혈증 예측 진단 성능에 대한 평가가 수행되었다. 한편, 선행 학습된 균혈증 예측 모델은, 스스로 라벨 값을 학습하는 오토 인코더 (AutoEncoder) 에 대하여 가중치 이니셜라이징 (weight initializing) 이 적용되었다.

[0177] 보다 구체적으로, 선행 학습된 본 발명의 균혈증 예측 모델 (pret) 의 경우, 균혈증 양성 또는 음성 예측의 정확도가 0.759, 민감도가 0.762로, 그렇지 않은 모델의 정확도 (0.729), 민감도 (0.75) 보다 높은 것으로 나타난다. 특히, 선행 학습된 본 발명의 균혈증 예측 모델 (pret) 의 진단 능력을 의미할 수 있는 AUC 값은 0.840으로, 그렇지 않은 모델의 AUC 값인 0.818보다 높은 것으로 나타난다.

[0178] 이상의 결과에 따르면, 본 발명의 다양한 실시예에 따른 균혈증 예측 모델은, 높은 민감도를 유지하면서 높은 AUC 값을 갖는, 오토 인코더에 대하여 가중치 이니셜라이징이 적용된 모델일 수 있다. 그러나 이에 제한되는 것은 아니다.

[0179] 도 5d 및 5e를 참조하면, 전술한 B 데이터 및 C 데이터에 기초하여 학습된 균혈증 예측 모델에 대한 레이어 사



이즈에 따른 평가 결과가 도시된다.

- [0180] 먼저, 도 5d를 참조하면, 64 개의 노드 128 개의 노드 및 64 개의 노드 각각으로 구성된 3 층의 복수의 레이어 (64-128-64) 로 구성된 예측 모델, 64 개의 노드 128 개의 노드, 128 개의 노드 및 64 개의 노드 각각으로 구성된 4 층의 복수의 레이어 (64-128-128-64) 로 구성된 예측 모델에 대하여 전술한 B 데이터를 이용한 성능 평가 결과가 도시된다.
- [0181] 이때, 64-128-64 예측 모델 및 64-128-128-64 예측 모델 각각은, 0.5 확률의 do (drop-out) 또는, bn (batch normalization) 이 적용되었다.
- [0182] 보다 구체적으로, bn을 적용한 64-128-64 예측 모델 및 64-128-128-64 예측 모델 각각에 대한 균혈증 예측의 정확도는 0.735, 0.738로 레이어 사이즈의 변화에도 유사한 값을 갖는 것으로 나타난다. 또한, bn을 적용한 64-128-64 예측 모델 및 64-128-128-64 예측 모델 각각에 대한 균혈증 예측에 대한 AUC 값은 0.847, 0.859로 레이어 사이즈의 변화에도 유사한 값을 갖는 것으로 나타난다.
- [0183] 한편, 도 5e를 더욱 참조하면, 128 개의 노드로 구성된 단일 레이어 (128) 로 구성된 예측 모델, 64 개의 노드 128 개의 노드 및 64 개의 노드 각각으로 구성된 3 층의 복수의 레이어 (64-128-64) 로 구성된 예측 모델, 64 개의 노드 128 개의 노드, 128 개의 노드 및 64 개의 노드 각각으로 구성된 4 층의 복수의 레이어 (64-128-128-64) 로 구성된 예측 모델에 대하여 전술한 C 데이터를 이용한 성능 평가 결과가 도시된다.
- [0184] 이때, 128 예측 모델에 대하여 0.5 확률의 do를 적용하였고, 64-128-64 예측 모델 및 64-128-128-64 예측 모델 각각에 대하여, 0.5 확률의 do 및/또는 bn이 적용되었다.
- [0185] 보다 구체적으로, 단일 레이어의 128 예측 모델의 경우 AUC 값은 0.751로 복수의 레이어를 갖는 64-128-64 예측 모델 및 64-128-128-64 예측 모델의 AUC 값보다 높은 것으로 나타난다.
- [0186] 이때, 64-128-64 예측 모델 및 64-128-128-64 예측 모델의 경우, 전술한 도 5d의 결과와 유사하게 bn을 적용했을 때의 균혈증 예측의 정확도, 민감도 및 AUC 값이 높은 것으로 나타난다.
- [0187] 이상의 결과에 따르면, 본 발명의 다양한 실시예에 따른 균혈증 예측 모델은, 레이어 사이즈의 변화에도 높은 민감도, 정확도 및 AUC 값을 갖는 것으로 나타난다. 이에, 본 발명의 다양한 실시예에 따른 균혈증 예측 모델은, 128 개의 노드로 구성된 단일 레이어를 갖거나, 64-128-64의 3층의 복수의 레이어를 갖거나, 64-128-128-64의 4층의 복수의 레이어를 갖고, bn이 적용된 모델일 수 있다. 그러나, 본 발명의 균혈증 예측 모델의 구조는 이에 제한되는 것은 아니다.
- [0188] 도 5f 및 5g를 참조하면, 전술한 B 데이터 및 C 데이터에 기초하여 학습된 스택킹 앙상블 (stacking ensemble) 모델에 대한 평가 결과가 도시된다.
- [0189] 먼저, 도 5f를 참조하면, 64 개의 노드 128 개의 노드 및 64 개의 노드 각각으로 구성된 3 층의 복수의 레이어 (64-128-64) 를 갖고, 스택킹 모델의 수가 10 개, 15 개 및 20 개인 앙상블 모델에 대한, 전술한 B 데이터를 이용한 성능 평가 결과가 도시된다.
- [0190] 보다 구체적으로, B 데이터 세트를 적용하였을 때, 스택킹 모델의 수가 15 개인 앙상블 모델에 대한 AUC 값이 0.887로, 스택킹 모델의 수가 10 개인 앙상블 모델의 AUC 값 (0.865), 스택킹 모델의 수가 20 개인 앙상블 모델의 AUC 값 (0.873) 보다 높은 것으로 나타난다.
- [0191] 한편, 도 5g를 함께 참조하면, 128 개의 노드로 구성된 단일 레이어를 갖고, 스택킹 모델의 수가 10 개, 15 개 및 20 개인 앙상블 모델에 대한, 전술한 C 데이터를 이용한 성능 평가 결과가 도시된다.
- [0192] 보다 구체적으로, C 데이터 세트를 적용하였을 때, 스택킹 모델의 수가 15 개 및 20 개인 앙상블 모델에 대한 AUC 값이 0.752로, 스택킹 모델의 수가 10 개인 앙상블 모델의 AUC 값 (0.750) 보다 높은 것으로 나타난다.
- [0193] 이상의 결과에 따르면, 본 발명의 다양한 실시예에 따른 균혈증 예측 모델은, 스택킹 모델의 수가 15 개 내지 20 개인 앙상블 모델일 수도 있다. 그러나, 본 발명의 균혈증 예측 모델의 구조는 이에 제한되는 것은 아니다.
- [0194] 도 5h 및 5i를 참조하면, 전술한 B 데이터 및 C 데이터에 기초하여 학습된, MLP 기반의 예측 모델 및 Xgboost 기반의 예측 모델에 대한 평가 결과가 도시된다.
- [0195] 먼저, 도 5h를 참조하면, 64 개의 노드 128 개의 노드 및 64 개의 노드 각각으로 구성된 3 층의 복수의 레이어 (64-128-64) 를 갖고, 스택킹 모델의 수가 15 개인 MLP 모델 및 gbtree 레이어로 구성된 Xgboost 모델에 대한,

전술한 B 데이터를 이용한 성능 평가 결과가 도시된다.

- [0196] 보다 구체적으로, B 데이터 세트를 적용하였을 때, Xgboost 모델 기반의 예측 모델의 AUC 값은 0.842, 민감도는 0.745로, MLP 기반의 예측 모델의 AUC 값 (0.887) 및 민감도 (0.801) 와 유사한 수준을 갖는 것으로 나타난다.
- [0197] 한편, 도 5i를 함께 참조하면, 128 개의 노드로 구성된 단일 레이어를 갖고, 스택킹 모델의 20 개인 MLP 모델 및 gbtree 레이어로 구성된 Xgboost 모델에 대한, 전술한 C 데이터를 이용한 성능 평가 결과가 도시된다.
- [0198] 보다 구체적으로, C 데이터 세트를 적용하였을 때, Xgboost 모델 기반의 예측 모델의 AUC 값은 0.745, 민감도는 0.635로, MLP 기반의 예측 모델의 AUC 값 (0.752) 및 민감도 (0.664) 와 유사한 수준을 갖는 것으로 나타난다.
- [0199] 이상의 결과에 따르면, 본 발명의 다양한 실시예에 따른 균혈증 예측 모델은, MLP 기반의 예측 모델 또는 Xgboost 기반의 예측 모델일 수 있다. 그러나, 본 발명의 균혈증 예측 모델은 이에 제한되지 않고 보다 다양한 예측 모델에 기초하여 균혈증 양성 또는 음성을 예측하도록 구성된 모델일 수 있다.
- [0200] 이상의 실시예 2의 결과로, 본 발명의 다양한 실시예에서 이용된 균혈증 예측 모델이 균혈증 발병 위험도를 높은 정확도로 예측하는 것을 확인할 수 있었다. 나아가, 높은 AUC 값을 갖는 것으로 나타난 결과에 따라 본 발명의 예측 모델은 우수한 진단 능력을 가지는 것으로 확인할 수 있었다.
- [0201] 이에, 본 발명은 균혈증에 대한 처치 시점을 앞당겨 치료에 대한 좋은 예후를 제공할 수 있다. 나아가, 본 발명은, 폐혈증과 상이한 균혈증에 대하여 높은 신뢰도 및 정확도로 발병의 위험도를 예측할 수 있는 효과가 있다.
- [0202] 특히, 본 발명은 개체로부터 빠르게 획득 가능한 임상 데이터에 기초한 균혈증 발병 위험도 예측 시스템을 제공함으로써, 개체에 대하여 균혈증 발병의 조기 진단을 제공할 수 있다.
- [0203] 이에, 본 발명은 균혈증 발병 개체의 생존률 증가, 합병증 예방, 및 치료 비용의 감소 효과를 제공할 수 있다.
- [0204] 또한, 본 발명은, 개체에 대하여 예측 모델에 의해 균혈증이 예측될 경우, 알람을 제공하여 균혈증 발병 위험도를 알리도록 구성됨에 따라, 보호자 또는 의료진은 중환자와 같은 지속적인 모니터링이 요구되는 개체에 대하여 보다 용이하게 균혈증 발병 고위험군을 인지할 수 있다.
- [0205] 본 발명의 여러 실시예들의 각각 특징들이 부분적으로 또는 전체적으로 서로 결합 또는 조합 가능하며, 당업자가 충분히 이해할 수 있듯이 기술적으로 다양한 연동 및 구동이 가능하며, 각 실시예들이 서로에 대하여 독립적으로 실시 가능할 수도 있고 연관 관계로 함께 실시 가능할 수도 있다.
- [0206] 이상 첨부된 도면을 참조하여 본 발명의 실시예들을 더욱 상세하게 설명하였으나, 본 발명은 반드시 이러한 실시예로 국한되는 것은 아니고, 본 발명의 기술사상을 벗어나지 않는 범위 내에서 다양하게 변형 실시될 수 있다. 따라서, 본 발명에 개시된 실시예들은 본 발명의 기술 사상을 한정하기 위한 것이 아니라 설명하기 위한 것이고, 이러한 실시예에 의하여 본 발명의 기술 사상의 범위가 한정되는 것은 아니다. 그러므로, 이상에서 기술한 실시예들은 모든 면에서 예시적인 것이며 한정적이 아닌 것으로 이해해야만 한다. 본 발명의 보호 범위는 아래의 청구범위에 의하여 해석되어야 하며, 그와 동등한 범위 내에 있는 모든 기술 사상은 본 발명의 권리범위에 포함되는 것으로 해석되어야 할 것이다.

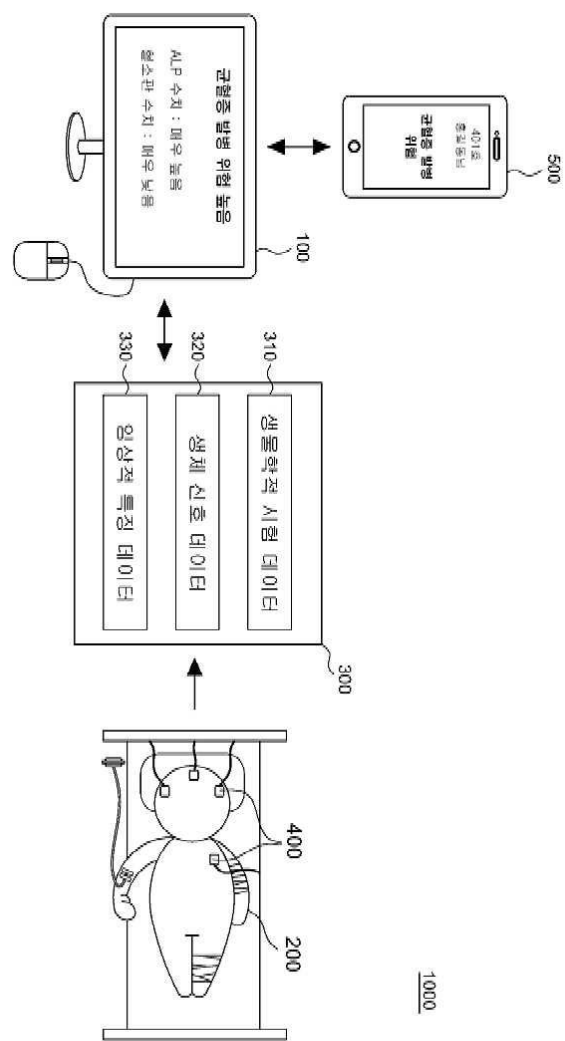
## 부호의 설명

- [0207] 100: 균혈증 발병 위험도 예측용 디바이스  
 110: 수신부  
 120: 입력부  
 130: 출력부  
 140: 저장부  
 150: 프로세서  
 200: 개체  
 300: 임상 데이터

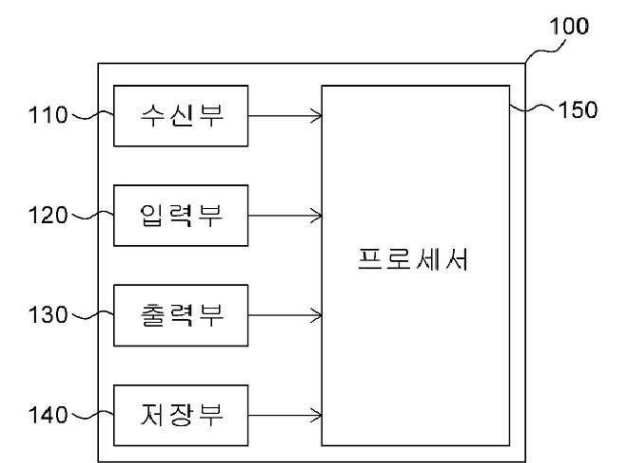
- 310: 생물학적 시험 데이터
- 320: 생체 신호 데이터
- 330: 임상적 특징 데이터
- 400: 의료적 처치 기구
- 500: 의료진 디바이스
- 1000: 위험도 예측 시스템

도면

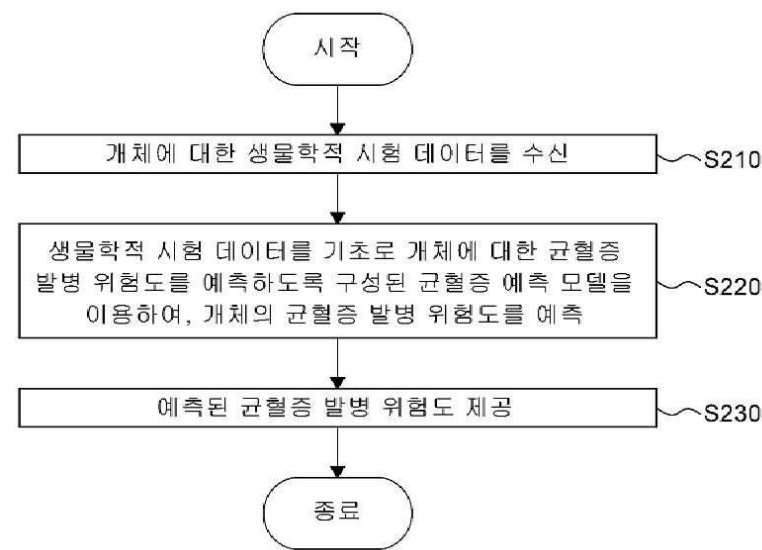
도면1a



도면1b



도면2



도면3a

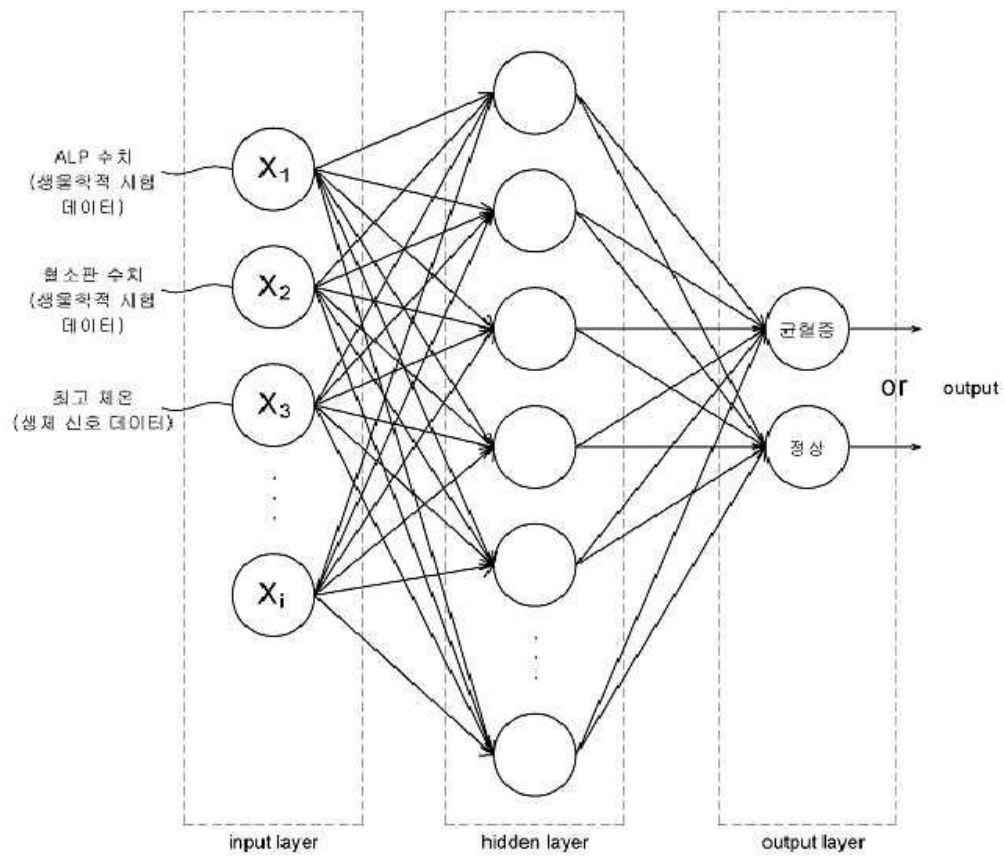
	균혈증 발병 개체	정상 개체	전체
학습 데이터	1260	1260	2520



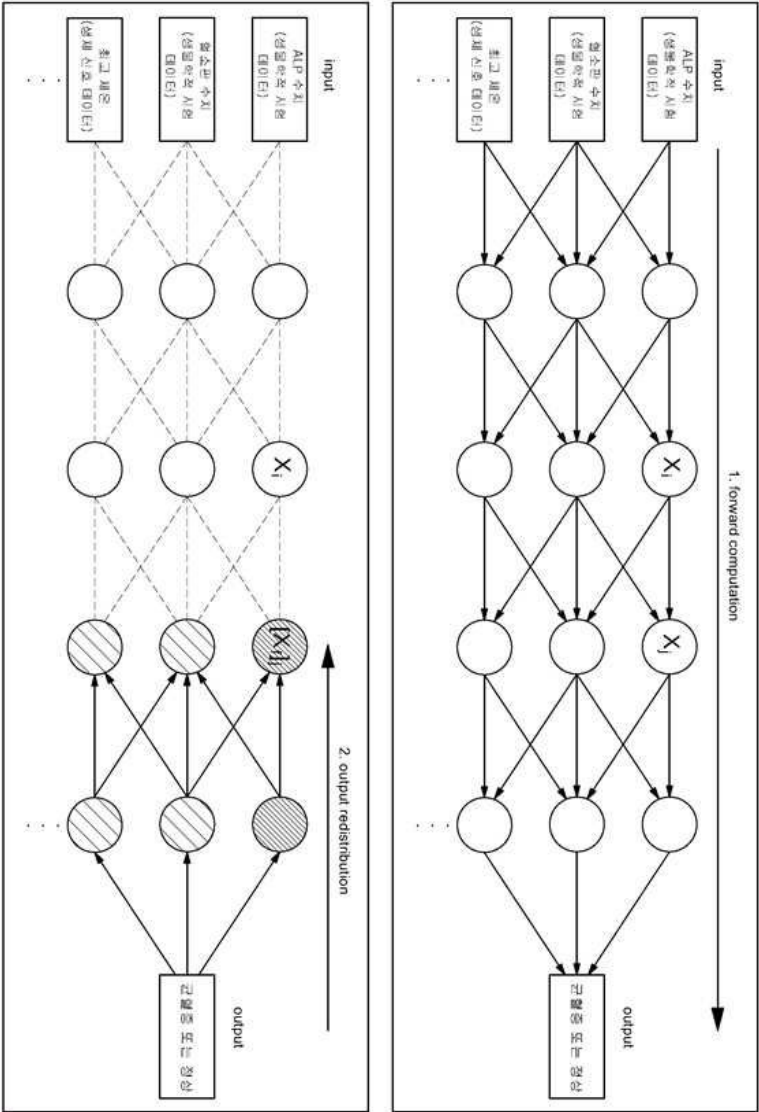
도면3b

	균혈증 (n=1,260)	정상 (n=1,260)	P-value
<b>생체 신호 데이터</b>			
SBP, mmHg	91.6 ± 23.2	99.4 ± 22.8	< 0.001
최저 체온, °C	36.1 ± 0.4	36.1 ± 0.3	0.578
최고 체온, °C	38.1 ± 0.9	37.6 ± 0.9	< 0.001
심박수, beats/min	115.7 ± 29.2	105.7 ± 28.5	< 0.001
호흡수, /min	25.5 ± 8.6	23.6 ± 8.9	< 0.001
<b>생물학적 시험 데이터</b>			
크레아틴, mg/dL	1.6 ± 1.5	1.2 ± 1.3	0.317
알부민, g/dL	2.8 ± 0.6	3.2 ± 0.7	< 0.001
CRP, mg/L	1.6 ± 1.5	1.2 ± 1.3	< 0.001
ALP, IU/L	191.3 ± 213.7	136.0 ± 141.0	< 0.001
WBC 최소 수치, /μL	10.3 ± 6.2	9.3 ± 5.5	< 0.001
WBC 최대 수치, /μL	14.8 ± 8.9	11.9 ± 7.3	< 0.001
혈소판 수, /μL	185.9 ± 134.8	218.7 ± 129.8	< 0.001
프로트롬빈 시간, sec	18.5 ± 8.9	15.9 ± 6.6	< 0.001
<b>임상적 특징 데이터</b>			
혈액 배양 기간, days	24.6 ± 75.3	10.3 ± 17.6	< 0.001
ICU 치료 여부, yes	410 (32.6%)	235 (18.7%)	< 0.001
중심 정맥 카테터 여부, yes	114 (9.1%)	0 (0%)	< 0.001
스테로이드 치료 여부, yes	477 (37.9)	429 (34.0)	0.045
항생제 치료 여부, yes	591 (46.9%)	526 (41.7%)	0.009
연령, years	64.4 ± 15.3	60.1 ± 17.3	< 0.001
성별, male	558 (44.3%)	533 (42.3%)	0.306

도면3c



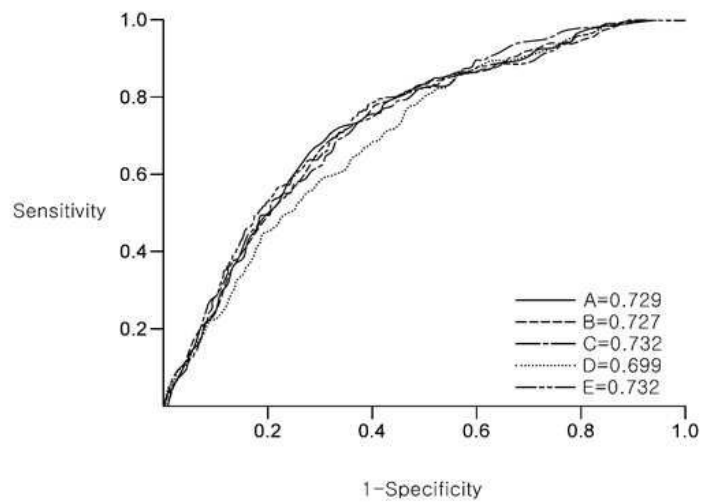
도면3d



도면4a

Model	Algorithms	AUC (95% CI)	Sensitivity	Specificity
종래의 예측모델	Naive Bayesian	0.7	-	-
예측 모델 (A)	MLP (128)	0.729 (0.712 - 0.728)	0.81	0.589
예측 모델 (B)	MLP (256)	0.727 (0.713 - 0.727)	0.81	0.532
예측 모델 (C)	RF	0.732 (0.722 - 0.733)	0.682	0.655
예측 모델 (D)	SVM	0.699 (0.687 - 0.700)	0.692	0.553
예측 모델 (E)	MLP (128)+ MLP (256)+ RF	0.732 (0.728 - 0.735)	0.793	0.571

도면4b



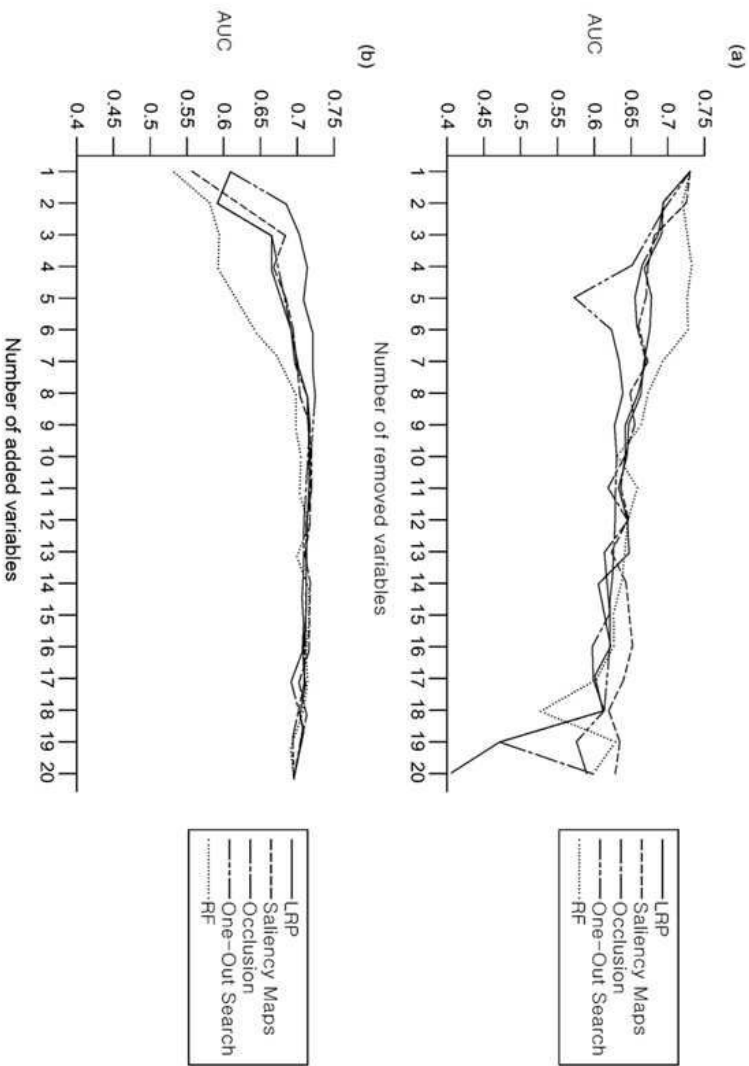
도면4c

Rank	One-out search	Gini Importance
1	ALP	연령
2	혈소판	프로트롬빈 시간
3	최고 체온	심박수
4	SBP	혈액 배양 기간
5	WBC 최소 수치	알부민
6	ICU 치료	ALP
7	CRP	혈소판
8	중심 정맥 카테터	최고 체온
9	프로트롬빈 시간	WBC 최고 수치
10	알부민	크레아틴
11	스테로이드 치료	SBP
12	성별	WBC 최소 수치
13	항생제 치료	호흡수
14	최저 체온	ICU 치료
15	혈액 배양 기간	스테로이드 치료
16	WBC 최고 수치	성별
17	심박수	최저 체온
18	호흡수	항생제 치료
19	연령	CRP
20	크레아틴	중심 정맥 카테터

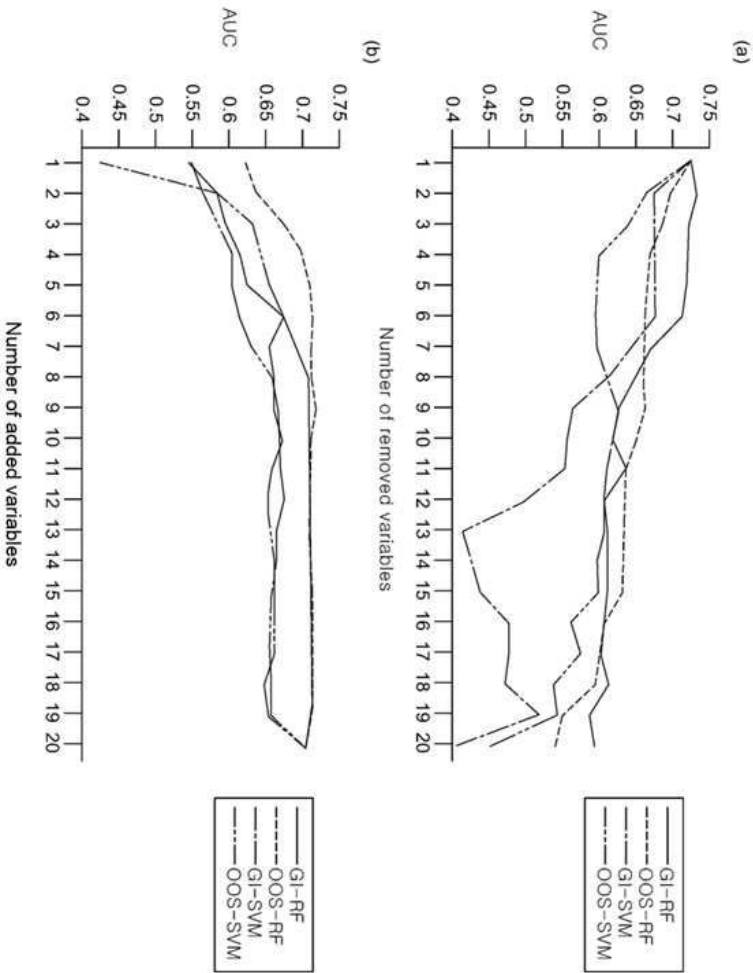
도면4d

Rank	Gradient-based methods			
	Layer-wise Relevance Propagation	Gradient Input	Integrated Gradients	Saliency Maps
1	ALP	ALP	ALP	혈액 배양 기간
2	연령	연령	연령	혈소판
3	혈소판	혈소판	혈소판	ALP
4	혈액 배양 기간	혈액 배양 기간	혈액 배양 기간	연령
5	심박수	CRP	심박수	ICU 치료
6	CRP	심박수	CRP	심박수
7	SBP	SBP	SBP	SBP
8	최고 체온	최고 체온	최고 체온	CRP
9	WBC 최소 수치	WBC 최소 수치	WBC 최소 수치	최고 체온
10	크레아틴	크레아틴	크레아틴	스테로이드 치료
11	최저 체온	WBC 최고 수치	최저 체온	WBC 최소 수치
12	WBC 최고 수치	호흡수	WBC 최고 수치	크레아틴
13	알부민	알부민	알부민	항생제 치료
14	프로트롬빈 시간	최저 체온	프로트롬빈 시간	성별
15	스테로이드 치료	프로트롬빈 시간	호흡수	중심 정맥 카테터
16	호흡수	스테로이드 치료	스테로이드 치료	WBC 최고 수치
17	항생제 치료	항생제 치료	항생제 치료	호흡수
18	ICU 치료	성별	ICU 치료	최저 체온
19	성별	ICU 치료	성별	알부민
20	중심 정맥 카테터	중심 정맥 카테터	중심 정맥 카테터	프로트롬빈 시간

도면4e



도면4f



도면5a

범주	특징 구성	비고
임상적 특징 데이터	연령, 성별	
생체 신호 데이터	SBP (systolic blood pressure), DBP (diastolic blood pressure), 최고 체온, 최저 체온, 심박수 및 호흡수	최소값, 최대값
생물학적 시험 데이터	알부민 수치, ALT (alanine aminotransferase) 수치, 트롬보 플라스틴 (thromboplastin) 의 응고 시간, AST (aspartate aminotransferase) 수치, CRP (C-reactive protein) 수치, 적혈구 침강 속도, FERR (ferritin) 수치, HMG (hemoglobin) 수치, PTINR (Prothrombin Time International Normalized Ratio), PTPER, PTSEC, 적혈구 수치, TBIL (Total bilirubin) 수치, TCO <sub>2</sub> (total carbon dioxide) 수치 및 백혈구 수치	

도면5b

분류	Train		Validation		Test	
	양성	음성	양성	음성	양성	음성
A	54,058	54,058	6,007	6,007	18,747	181,816
B	55,168	55,168	7,881	7,881	15,763	116,293
C	55,176	55,233	7,885	7,907	15,751	116,299



도면5c

DB	데이터 구조	구분	in	lo	hi	lo	AUC	accuracy	TPR {recall}	TNR {specificity}	NPV	FPR
		Per	88.170	28.123	3.744	12.019	0.840	0.759	0.762	0.758	0.989	0.242
A	128-64-25-128		84.434	31.859	3.945	11.818	0.818	0.729	0.75	0.725	0.985	0.274

도면5d

DB	데이터 구조	구분	in	to	ln	ld	AUC	Accuracy	TPR (recall)	TNR (specificity)	NPV	FPR
		do 0.5	71.390	44.913	4.350	11.413	0.732	0.627	0.724	0.614	0.943	0.388
B	64-128-64	bn	84.776	31.517	3.450	12.313	0.847	0.735	0.781	0.729	0.961	0.271
		do 0.5	75.746	40.547	4.902	10.861	0.733	0.666	0.699	0.651	0.939	0.349
B	64-128-128-64	bn	84.786	31.507	3.055	12.708	0.859	0.738	0.806	0.729	0.965	0.271

도면5e

DB	레이아웃 구조	구분	tn	fp	fn	tp	AUC	Accuracy	TPR (recall)	TNR (specificity)	NPV	FPR
	128	do 0.5	84.302	31.997	5.609	10.142	0.751	0.715	0.644	0.725	0.938	0.275
C	64-128-64	do 0.5	84.029	32.270	5.633	10.118	0.748	0.713	0.642	0.723	0.937	0.277
		bn	77.673	38.626	5.863	9.868	0.702	0.683	0.626	0.688	0.93	0.332
		bn, do 0.5	84.029	30.304	5.970	9.781	0.749	0.721	0.621	0.735	0.934	0.265
		do 0.5	83.841	32.458	5.683	10.068	0.745	0.711	0.639	0.721	0.937	0.279
	64-128-128-64	bn	78.482	37.817	5.468	10.283	0.719	0.672	0.653	0.675	0.935	0.325
		bn, do 0.5	83.761	32.588	5.641	10.110	0.748	0.711	0.642	0.72	0.937	0.28

도면5f

DB	데이터 구조	구분	tn	fp	fn	tp	AUC	Accuracy	TPR (recall)	TNR (specificity)	NPV	FPR
		n=10	91.344	24.949	3.717	12.046	0.855	0.783	0.764	0.785	0.981	0.215
		n=15	87.895	26.398	3.143	12.620	0.887	0.761	0.801	0.756	0.985	0.244
		n=20	92.180	24.133	3.595	12.177	0.873	0.79	0.773	0.792	0.983	0.208
		B4-128-B4										

도면5g

DB	레이더 구조	구분	in	fo	fn	fp	AUC	Accuracy	TPR (focal)	TNR (specificity)	NPV	FPR
		n=10	81.895	34.404	5.336	10.415	0.75	0.899	0.861	0.704	0.939	0.286
C	128	n=15	80.205	36.094	5.075	10.676	0.752	0.888	0.878	0.68	0.94	0.31
		n=20	81.459	34.840	5.298	10.453	0.752	0.896	0.864	0.7	0.939	0.3

도면5h

CG	레이어 구조	구분	in	fo	fn	to	AUC	Accuracy	TPR (recall)	TNR (specificity)	NPV	FPR
	84-128-64 bytes	MLP XGBoost	67.895 89.398	28.398 28.895	3.143 4.015	12.620 11.748	0.887 0.842	0.761 0.766	0.801 0.745	0.766 0.769	0.985 0.957	0.244 0.231

도면5i

DB	240이 구조	구분	tn	tp	fn	fp	AUC	Accuracy	TFR (recall)	TNR (specificity)	NPV	FPR
	128	MLP	81.459	34.840	5.298	10.453	0.792	0.898	0.664	0.7	0.939	0.3
C	gcnne	XGBoost	84.064	32.253	5.743	10.006	0.745	0.712	0.635	0.723	0.896	0.277